

Análise de Seleção de Características e Otimização de Hiperparâmetros via Algoritmos Evolutivos

1st Agnelo Pereira Lima Júnior
Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Espírito Santo (UFES)
Vitória, ES, Brazil
email address or ORCID

Abstract—Este trabalho detalha a implementação e os resultados de um sistema de otimização para classificação e regressão de dados, utilizando Algoritmos Evolutivos (AE) para realizar a seleção de características e o ajuste de hiperparâmetros de Redes Neurais Feedforward. Foram avaliados os algoritmos Genético (GA) e de Evolução Diferencial (DE) em conjunto com dois métodos de treino: Backpropagation e Extreme Learning Machine (ELM). Os experimentos, conduzidos nos datasets Wine, Hepatitis e Diabetes, demonstraram que a abordagem com Backpropagation foi robusta e eficaz, alcançando acurácias de até 85% para classificação e um R^2 de 0.48 para regressão. Em contrapartida, a implementação com ELM mostrou-se instável.

Index Terms—Seleção de Características, Algoritmo Genético, Evolução Diferencial, Rede Neural, Backpropagation, Regressão.

I. INTRODUÇÃO

A crescente dimensionalidade dos dados modernos tornou a seleção de características (feature selection) uma etapa crucial em projetos de Machine Learning. A remoção de atributos irrelevantes ou redundantes não só melhora a performance preditiva dos modelos, mas também reduz a complexidade computacional. Esta tarefa explora o uso de Algoritmos Evolutivos como uma técnica de busca heurística para encontrar o subconjunto ótimo de características e, simultaneamente, otimizar os hiperparâmetros de uma Rede Neural classificadora e regressora.

II. METODOLOGIA

O pipeline experimental foi construído em um notebook Jupyter [?] e seguiu os seguintes passos:

A. Representação do Indivíduo e Função de Fitness

Cada solução candidata (indivíduo) foi representada por um vetor que codifica uma máscara binária para as características e os hiperparâmetros da rede. A qualidade de cada indivíduo foi avaliada por uma função de fitness baseada na fórmula do artigo de referência, que pondera a taxa de erro e a redução de características:

$$\text{Fitness} = \rho \cdot \text{Err}(D) + (1 - \rho) \cdot \frac{F}{T} \quad (1)$$

onde ρ foi definido como 0.9. Para classificação, o erro foi a taxa de erro de acurácia, e para regressão, o Erro Quadrático Médio (MSE).

Identify applicable funding agency here. If none, delete this.

B. Algoritmos e Classificadores

Foram utilizados o GA e o DE da biblioteca `scikit-opt`. Para classificação, foram testados os métodos de treino Backpropagation (`MLPClassifier`) e ELM (`hpelem`). Para regressão, utilizou-se o Backpropagation (`MLPRegressor`).

III. CONFIGURAÇÃO EXPERIMENTAL

A. Ambiente de Execução

O treinamento e a otimização foram realizados em um ambiente com a seguinte configuração:

- **Processador (CPU):** 12th Gen Intel(R) Core(TM) i7-1265U 1.80 GHz
- **Memória RAM:** 16,0 GB
- **Sistema Operacional:** Sistema operacional de 64 bits, processador baseado em x64

B. Parâmetros de Treinamento

Os principais parâmetros utilizados nos experimentos foram [1]:

- **Tamanho da População:** 30 (Classificação) / 30 (Regressão)
- **Número de Gerações:** 50 (Classificação) / 40 (Regressão)
- **Neurônios por Camada:** Otimizado no intervalo [5, 50] para classificação e [5, 100] para regressão.
- **Taxa de Aprendizado:** Otimizado no intervalo [0.001, 0.1].
- **Validação Cruzada (avaliação):** 5-fold (Classificação) / 3-fold (Regressão)

IV. RESULTADOS E ANÁLISE

A. Dataset: Wine (Classificação)

A abordagem com Backpropagation mostrou-se eficaz. O algoritmo campeão foi o DE, resultando em uma acurácia de **85%** no conjunto de teste. O método ELM, contudo, foi instável, alcançando apenas 63% de acurácia.

B. Dataset: Hepatitis (Classificação)

Para este dataset, o DE com Backpropagation também foi superior, alcançando ****81%** de acurácia**. O método ELM falhou, com uma acurácia de apenas 17%.

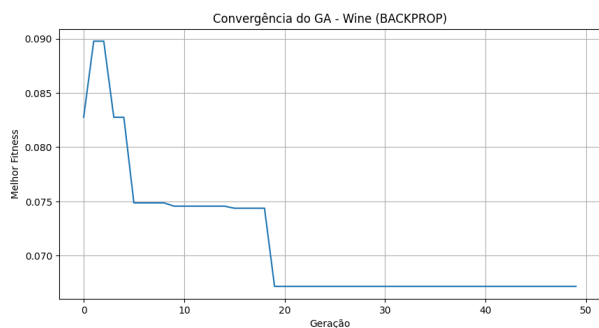


Fig. 1. Convergência do GA (Wine - Backpropagation).

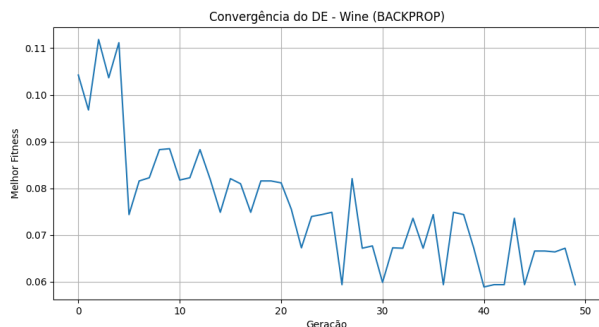


Fig. 2. Convergência do DE (Wine - Backpropagation).

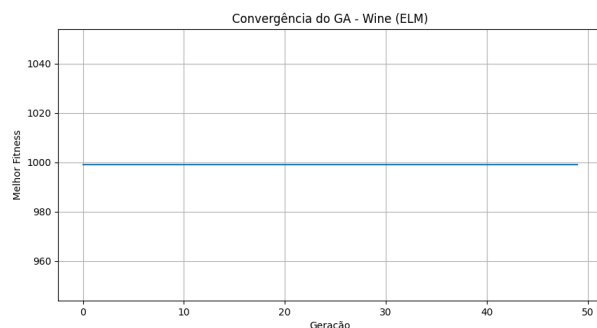


Fig. 3. Convergência do GA (Hepatitis - Backpropagation).

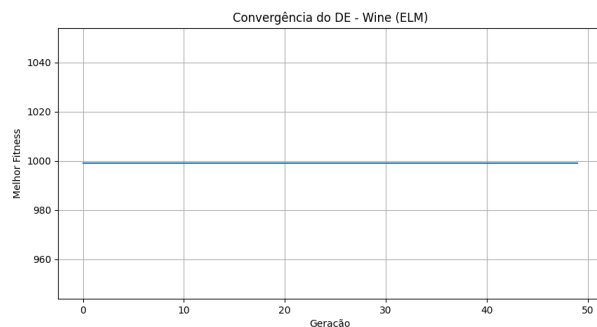


Fig. 4. Convergência do DE (Hepatitis - Backpropagation).

C. Dataset: Diabetes (Regressão)

A metodologia foi adaptada para regressão. O GA foi o algoritmo campeão. As métricas de eficácia no teste foram um **MSE de 2811.06** e um **R² de 0.4793**. O artigo de referência não aborda este dataset, impossibilitando uma comparação direta.

D. Análise Comparativa com o Artigo

TABLE I
COMPARAÇÃO DE ACURÁCIA (CLASSIFICAÇÃO)

Dataset	Resultado Obtido (NN)	Artigo (KNN)
Wine	85.0%	95.7%
Hepatitis	81.0%	87.5%

Os resultados de acurácia foram inferiores aos do artigo. A diferença pode ser atribuída ao uso de classificadores distintos (Rede Neural vs. KNN), indicando que o KNN pode ser mais adequado para estes datasets específicos no contexto da metodologia do artigo.

V. CONCLUSÃO

O framework implementado foi bem-sucedido em aplicar Algoritmos Evolutivos para a otimização de Redes Neurais em tarefas de classificação e regressão. A abordagem com Backpropagation mostrou-se robusta e eficaz, enquanto o ELM se mostrou instável. A metodologia foi validada como uma ferramenta funcional para a seleção de características e otimização de hiperparâmetros.

REFERENCES

- [1] Código-fonte e resultados gerados a partir do notebook Jupyter *feature_select_ga_de.ipynb*, conforme os experimentos realizados.

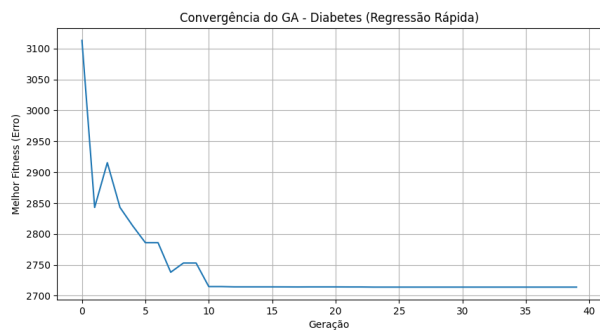


Fig. 5. Convergência do GA (Diabetes - Regressão).

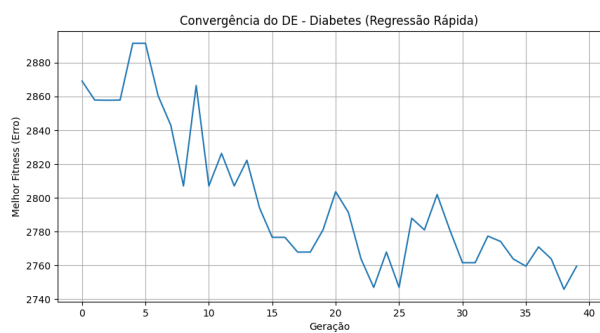


Fig. 6. Convergência do DE (Diabetes - Regressão).