Seleção Incremental de Atributos em Classificador Fuzzy Evolutivo com Agrupamento Gaussiano Multivariado

Patrick Silva Menezes
Departamento de Computação
CEFET-MG
Divinópolis, MG, Brasil
000-0005-2095-6418

Fernanda P. S. Rodrigues
Modelagem Matemática e Computacional
CEFET-MG
Belo Horizonte, MG, Brasil
0000-0001-9983-0765

Emmanuel Tavares
Modelagem Matemática e Computacional
CEFET-MG
Belo Horizonte, MG, Brasil
0000-0002-9253-0525

Michel Pires da Silva
Departamento de Computação
CEFET-MG
Divinópolis, MG, Brasil
0000-0002-1343-1331

Alisson Marques da Silva
Modelagem Matemática e Computacional
CEFET-MG
Belo Horizonte, MG, Brasil
0000-0002-1023-6514

Daniel Leite
Departamento de Computação
Universidade de Paderborn
Paderborn, Alemanha
0000-0003-3135-2933

***Resumo*—Este trabalho apresenta um novo classificador fuzzy evolutivo que incorpora, em seu processo de aprendizado, um método de seleção de atributos baseado na razão entre as médias amostrais dos atributos. A estrutura do classificador pode se expandir, contrair ou ser atualizada por meio de um algoritmo de agrupamento com aprendizado participativo gaussiano multivariado, aliado a uma estratégia de procrastinação. Essa abordagem permite a formação de grupos inativos — não imediatamente associados a regras — que podem ser ativados conforme novas amostras se tornam disponíveis, conferindo maior flexibilidade ao modelo. Os resultados experimentais indicam que o classificador proposto alcança desempenho superior ou comparável ao dos principais classificadores evolutivos do estado da arte, configurando-se como uma alternativa promissora para aplicações em cenários de aprendizado incremental, em tarefas de classificação binária e multiclasse.**

***Index Terms*—Sistemas *Fuzzy* Evolutivo, Seleção Incremental de Atributos, Classificação.**

I. Introdução

Nos últimos anos, o volume de dados gerado tem crescido de forma exponencial, impulsionado por avanços em tecnologias como sensores inteligentes, Internet das Coisas e sistemas de informação [1]. Esse aumento substancial impõe desafios significativos à análise e ao processamento de grandes volumes de informação, especialmente no que se refere aos algoritmos de inteligência artificial e aprendizagem de máquina [2].

Para lidar com esses desafios, métodos avançados de representação, extração de características e otimização têm sido intensivamente explorados, principalmente devido ao crescimento do volume e da dimensionalidade dos dados. A alta variabilidade das amostras impõe a necessidade de estratégias que mitiguem redundâncias e proporcionem maior robustez e generalização aos modelos [3].

Adicionalmente, a natureza do aprendizado desempenha um papel crucial na escolha dos algoritmos apropriados. Nesse sentido, os classificadores podem ser categorizados em duas abordagens principais: *offline* e *online*. Classificadores *offline* operam em dados estáticos, exigindo reprocessamento completo para incorporar atualizações. Em contraste, os classificadores *online* permitem aprendizado contínuo, ajustando-se dinamicamente a novas amostras. Nesse último caso, modelos incrementais ajustam seus parâmetros de forma progressiva, sem alterar sua estrutura, enquanto modelos evolutivos se reconfiguram conforme padrões emergentes, oferecendo maior adaptabilidade em cenários dinâmicos [4].

Os sistemas evolutivos têm se destacado como uma abordagem eficaz para modelar ambientes dinâmicos, oferecendo soluções para a crescente complexidade e variabilidade dos dados. Em especial, os sistemas *fuzzy* evolutivos combinam a flexibilidade dos modelos evolutivos com a clareza e adaptabilidade dos sistemas baseados em lógica *fuzzy* [5], [6]. Esses modelos possibilitam a atualização contínua das regras *fuzzy*, aprimorando a representação da incerteza e aumentando a transparência na tomada de decisão. A estrutura das regras *fuzzy*, formulada por expressões no formato SE-ENTÃO, contribui de forma significativa para a compreensão do sistema, tornando-o ideal para aplicações que exigem tanto interpretabilidade quanto flexibilidade [7].

A seleção de atributos é uma etapa fundamental para a eficiência dos modelos [8]. A remoção de atributos irrelevantes ou redundantes pode reduzir significativamente o tempo de treinamento, mitigar os efeitos da maldição da dimensionalidade e melhorar a capacidade de generalização do classificador, minimizando o risco de *overfitting* [4]. A incorporação de mecanismos de seleção incremental de atributos em sistemas evolutivos é, portanto, uma estratégia promissora para lidar com conjuntos de dados de alta dimensionalidade de maneira eficiente e adaptativa [9].

A partir do contexto exposto, este trabalho propõe um novo classificador *fuzzy* evolutivo com seleção incremental de atributos, aplicável a tarefas de classificação binária e multiclasse. A proposta baseia-se em um método denominado MRFS (*Mean Ratio Feature Selection*) [10], que utiliza a razão entre médias amostrais para identificar, de forma dinâmica, os atributos mais representativos ao longo do tempo. A estrutura do classificador é construída de forma incremental, amostra a amostra, por meio de um algoritmo de agrupamento com aprendizado participativo gaussiano multivariado, combinado a uma estratégia de procrastinação. Essa estratégia permite a criação, exclusão, mesclagem e atualização de grupos e regras, além de possibilitar a geração de grupos inativos — os quais não são imediatamente associados a regras, mas podem ser ativados posteriormente, à medida que novas amostras se tornam disponíveis.

O restante deste artigo está organizado da seguinte forma. Na Seção II, são apresentados o classificador proposto e seus procedimentos de aprendizagem. A Seção III descreve os experimentos computacionais realizados, bem como os resultados comparativos obtidos com classificadores evolutivos alternativos, considerando tarefas de classificação binária e multiclasse em quatro conjuntos de dados amplamente utilizados na literatura. Por fim, a Seção IV encerra o estudo, recapitulando as principais contribuições e sugerindo direções para pesquisas futuras.

II. Classificador Proposto

O algoritmo do classificador proposto calcula a saída simultaneamente à atualização de sua estrutura e de seus parâmetros a cada nova amostra. A atualização estrutural é realizada por meio da inclusão, exclusão, mesclagem e atualização de grupos e regras, além da adição ou remoção de atributos. Os métodos de seleção de atributos e de criação e ativação de grupos são acionados sempre que uma amostra é classificada incorretamente, enquanto o método de atualização de grupos é aplicado quando a amostra é classificada corretamente. Por outro lado, os procedimentos de exclusão e mesclagem de grupos são executados a cada nova amostra, independentemente do acerto da classificação.

As principais etapas do fluxo do classificador proposto são sumarizadas a seguir:

1) **Definir hiperparâmetros e atributos:** etapa executada apenas uma vez para configurar os hiperparâmetros do classificador e definir os atributos iniciais que serão considerados no processo de aprendizado.
2) **Ler amostra:** se for a primeira amostra, cria-se o primeiro grupo, a classe e a regra correspondente. Caso contrário, passa-se para a próxima etapa.
3) **Calcular grau de pertinência:** calcular o grau de pertinência da amostra em relação a todos os grupos ativos e identificar o índice do grupo mais ativo.
4) **Obter a saída do classificador.**
5) **Verificar se a classificação está correta:** se estiver correta, atualizar o centro do grupo mais ativo e passar para a Etapa 9. Caso contrário, seguir para a Etapa 6.
6) **Verificar se a classe da amostra existe:** se a classe ainda não existir, criar um novo grupo, classe e regra, e passar para a Etapa 9. Caso a classe já exista, seguir para a próxima etapa.
7) **Executar o procedimento de seleção de atributos:** verificar se um novo atributo deve ser incluído, se algum atributo deve ser excluído ou se o conjunto atual de atributos deve ser mantido.
8) **Executar o procedimento de criação ou ativação de grupos inativos:** verificar se um grupo inativo deve ser ativado, se um novo grupo inativo deve ser criado ou se a estrutura atual de grupos deve ser mantida.
9) **Executar o procedimento de mesclagem de grupos:** verificar se dois grupos devem ser mesclados ou se a estrutura atual de grupos deve ser mantida.
10) **Executar o procedimento de exclusão de grupos:** verificar se algum grupo deve ser excluído ou se a estrutura atual de grupos deve ser mantida. Em seguida, retornar à Etapa 2.

A. Inicialização

A estrutura do classificador é iniciada sem conhecimento prévio e, a partir da primeira amostra recebida, são criados o primeiro grupo, a primeira classe e a primeira regra. No algoritmo proposto, o número de grupos, regras e classes é determinado com base nas amostras x^t, representadas por $[x_1^t \dots x_j^t \dots x_m^t]^T$, em que t indexa a amostra atual, j representa os atributos de entrada, e m corresponde ao número de atributos de entrada.

Os grupos são inicializados centrados no valor da amostra atual, conforme:

$$\mu_j^i \leftarrow x_j^t, \quad j = 1 \dots m, \tag{1}$$

em que μ é o centro do grupo e i indexa o grupo.

Ao criar um grupo i são incrementados em um: o número de grupos criados c; o número de grupos ativos σ; o número de classes k e; o número de amostras do grupo s_i. Em seguida, a matriz de dispersão Ψ do grupo i é inicializada por:

$$\Psi_i \leftarrow \Psi_{init}, \tag{2}$$

na qual Ψ é uma matriz de dimensão m x m, inicializada com uma matriz identidade ajustada Ψ_{init} com valores diagonais entre 10^{-1} a 10^{-4}. Em seguida, o vetor de grupos ativos Ω é atualizado na posição σ com o índice do grupo recém-criado:

$$\Omega_\sigma \leftarrow c. \tag{3}$$

A classe do grupo é definida quando a saída desejada y^t é disponibilizada. Então, o vetor de classes é atualizado com a classe da amostra:

$$\Theta_k \leftarrow y^t. \tag{4}$$

Após a criação do grupo, a regra correspondente é gerada. O antecedente da regra é representado por uma função de

pertinência gaussiana multivariada, cujo valor modal é determinado pelo centro do grupo, enquanto a dispersão é calculada a partir da matriz de dispersão. A função de pertinência gaussiana multivariada pode ser expressa por:

$$f(x^t) = e^{-\frac{1}{2}(x^t-\mu^i)(\Psi_i)^{-1}(x^t-\mu^i)^T}. \quad (5)$$

O consequente da regra é determinado pela classe da amostra. As regras *fuzzy* do modelo possuem a seguinte forma:

$$\begin{cases} R_1: & \text{Se } x^t \text{ é } F_1 \text{ Então } \hat{y}_1^t = \Theta_1^t, \\ & \vdots \\ R_i: & \text{Se } x^t \text{ é } F_i \text{ Então } \hat{y}_i^t = \Theta_k^t, \end{cases}$$

em que R_i é a i-ésima regra, x^t é a amostra atual, F_i é o antecedente representado pela i-ésima função de pertinência gaussiana multivariável, $\hat{y}_i^t$ é o consequente da i-ésima regra. Por fim, o contador de grupos Φ da classe l é incrementado em um e o tempo de inatividade do grupo ϱ_i é definido como 0.

O Algoritmo 1 sintetiza o processo de inicialização, criação de grupos, classes e regras.

Algoritmo 1. Criação grupos, classes e regras.

Inicializar centro μ^i - (1)
Atualizar c, σ, k e s_i
Inicializar Ψ_i - (2)
Atualizar $\Omega\sigma$ - (3)
Atualizar Θ_k - (4)
Criar regra R_i
Atualizar Φ_l e ϱ_i

B. Cálculo da Saída

Para obter a saída do classificador, encontra-se o grau de disparo do antecedente, que é normalizado por meio do cálculo da distância de Mahalanobis, a cada nova amostra x^t, sendo seu cálculo realizado por:

$$\tau_{R^i} = \frac{e^{[(x^t-\mu^i)(\Psi^i)^{-1}(x^t-\mu^i)^T]}}{\sum_{i=1}^{\sigma} e[(x^t-\mu^i)(\Psi^i)^{-1}(x^t-\mu^i)^T]}, \quad (6)$$

sendo τ o grau de disparo do antecedente normalizado, i o indexador das regras, Ψ é a matriz de dispersão e σ é o número de regras. Após definir o grau de disparo normalizado do antecedente, identifica-se o índice da regra com o maior grau de ativação, denotado por i^+:

$$i^+ = arg\ max_{i=1,2,...,p^t}\ \tau_{R^i}. \quad (7)$$

O consequente da regra mais ativa, $\hat{y}^{i^+} \in \{1, ..., k\}$, é fornecido como uma previsão para x^t.

O procedimento para se obter a saída do classificador é ilustrado no Algoritmo 2.

Algoritmo 2. Cálculo da saída.

// Para cada regra
for *i = 1 to p* **do**
  Calcular o grau de ativação τ_{R^i} - (6)
Encontrar o índice da regra mais ativa i^+ - (7)
Obter a saída $\hat{y}$

C. Criação e Atualização de Grupos e Regras

A atualização e a criação de grupos são realizadas por um algoritmo de agrupamento baseado em aprendizado participativo multivariado gaussiano, em conjunto com um mecanismo de procrastinação. O algoritmo de agrupamento emprega uma medida de compatibilidade [11], a qual avalia o grau de compatibilidade de uma nova amostra ao conhecimento existente, representado pela estrutura atual dos grupos. Essa medida é utilizada para ajustar dinamicamente a taxa de aprendizado, reduzindo-a para entradas atípicas (*outliers*) e aumentando-a para entradas compatíveis com o conhecimento atual [12].

O mecanismo de procrastinação, por sua vez, é projetado para lidar eficientemente com *outliers*. Nesse contexto, os grupos são inicialmente criados em estado inativo, isto é, sem associação a qualquer regra e sem influência na saída do classificador. A ativação de um grupo ocorre apenas quando o número de amostras acumuladas é suficiente para exceder um limiar de procrastinação predefinido, garantindo maior robustez ao processo de modelagem.

O método de agrupamento proposto é capaz de: (i) criar um novo grupo ativo; (ii) atualizar um grupo ativo existente; (iii) criar um novo grupo inativo; (iv) atualizar um grupo inativo; e (v) ativar um grupo inativo. A execução desses procedimentos é orientada pela saída do classificador, conforme detalhado a seguir:

- **Caso 1: Se** $\hat{y}^t = y^t$, ou seja, a amostra x^t foi classificada corretamente, **então** o centro do grupo mais ativo, indexado por i^+, é atualizado conforme:

$$\mu^{i^+} = \mu^{i^+} + \alpha\gamma^{i^+}(x^t - \mu^{i^+}), \quad (8)$$

em que α representa a taxa de aprendizado e γ_i^t a medida de compatibilidade, com valores variando entre $[0, 1]$, calculada por:

$$\gamma^{i^+} = e^{-\frac{1}{2}[(x^t-\mu^{i^+})(\Psi^{i^+})^{-1}(x^t-\mu^{i^+})^T]}. \quad (9)$$

Em seguida, a matriz de dispersão do grupo mais ativo Ψ_{i^+} é atualizada por:

$$\Psi^{i^+} = (1 - \alpha\gamma^{i^+})(\Psi^{i^+} - \alpha\gamma^{i^+}(x^t - \mu^{i^+})(x^t - \mu^{i^+})^T). \quad (10)$$

O Algoritmo 3 ilustra o procedimento de atualização do centro mais ativo.

Algoritmo 3. Atualização do centro do grupo mais ativo.

//Dado o índice do cluster mais ativo i^+ para a amostra x^t
Calcular γ^{i^+} - (9)
Atualizar μ^{i^+} - (8)
Atualizar Ψ^{i^+} - (10)

- **Caso 2: Se** $\hat{y}^t \neq y^t$, ou seja, a classificação foi incorreta e a classe y^t da amostra não existe no vetor de classes Θ_l^t $(l = 1 \dots k^t)$, **então** um novo grupo ativo, uma nova classe e uma nova regra são criados. Este procedimento é igual àquele descrito na Seção II-A (Algoritmo 1), portanto, não será detalhado novamente.
- **Caso 3: Se** $\hat{y}^t \neq y^t$, ou seja, a classificação foi incorreta e a classe y^t existe no vetor de classes dos grupos ativos $(\exists y^t = \Theta_l^t)$, mas não no vetor de classes dos grupos inativos $(\neg\exists y^t = \bar{\Theta}_l^t)$, **então** um novo grupo inativo é criado. Em seguida, os contadores de grupos inativos criados $\bar{c}$, de classes $\bar{k}$, de grupos inativos $\bar{\sigma}$ e de amostras do grupo inativo $\bar{s}$ são incrementados em um. Depois, a matriz de dispersão do grupo inativo é inicializada (2) e o vetor de grupos inativos $\bar{\Omega}$ é atualizado, indexado por $\bar{\sigma}$:

$$\bar{\Omega}_{\bar{\sigma}} \leftarrow \mu. \tag{11}$$

Em seguida, o vetor de classes dos grupos inativos é atualizado:

$$\bar{\Theta}_{\bar{k}} \leftarrow y^t. \tag{12}$$

Por fim, atualiza-se o contador de grupos $\bar{\Phi}$ inativos da classe l e o tempo de inatividade do grupo ϱ_i. O Algoritmo 4 sumariza o procedimento para criação de grupos inativos.

Algoritmo 4. Criação de grupo inativo.

Inicializar centro μ^i - (1)
Atualizar $\bar{c}$, $\bar{k}$, $\bar{\sigma}$ e $\bar{s}_i$
Inicializar Ψ_i - (2)
Atualizar $\bar{\Omega}_{\bar{\sigma}}$ - (11)
Atualizar $\bar{\Theta}_{\bar{k}}$ - (12)
Atualizar $\bar{\Phi}_l$ e ϱ_i

- **Caso 4: Se** $\hat{y}^t \neq y^t$, ou seja, a classificação foi incorreta e a classe da amostra y^t existe tanto nos grupos ativos $(\exists y^t = \Theta_l^t)$ quanto nos inativos $(\exists y^t = \bar{\Theta}_l^t)$, **então** verifica-se se um grupo inativo deve ser ativado ou atualizado. Primeiramente, é calculado o grau de ativação normalizado usando a equação (6) para os grupos inativos que representam a mesma classe da amostra. Em seguida, encontra-se o grupo inativo mais compatível com a amostra atual, e o seu índice i^+ é obtido usando (7). Em seguida, verifica-se se o grupo inativo indexado por i^+ deve ser atualizado ou ativado.
  Se o número de amostras $\bar{s}^{i^+}$ do grupo inativo indexado por i^+, for menor que o limiar de procrastinação η, então o número de amostras do grupo i^+ é incrementado em um, $\bar{s}^{i^+} = \bar{s}^{i^+} + 1$. Depois, o centro do grupo inativo é atualizado por (8) utilizando a medida de compatibilidade (9). Por fim, a matriz de dispersão do grupo inativo i^+ é atualizada (10).
  Por outro lado, caso o número de amostras $\bar{s}^{i^+}$ seja igual ou maior que o limiar de procrastinação η, o grupo inativo indexado por i^+ é ativado. Então, o número de grupos ativos σ e o vetor de grupos ativos Ω são atualizados. Em seguida, uma nova regra é criada, cujo consequente é obtido pela classe do grupo recém ativado:

$$\hat{y}^{i^+} \leftarrow \bar{\Theta}_l. \tag{13}$$

Por fim, o número de grupos da classe l, Φ_l, é incrementado em uma unidade, e o tempo de inatividade do grupo é definido como zero, ou seja, $\varrho_i = 0$.

O Algoritmo 5 sumariza o procedimento para atualização e ativação de grupo inativo.

Algoritmo 5. Atualização e ativação de grupo inativo.

Calcular o grau de ativação - (6)
Encontrar o índice do grupo mais ativo $\bar{i}^+$ - (7)
if $\bar{s}^{i^+} < \eta$ **then**
  Atualizar $\bar{s}^{i^+}$
  Atualizar γ^{i^+} - (9)
  Atualizar μ^{i^+} - (8)
  Atualizar Ψ^{i^+} - (10)
else
  Ativar grupo inativo indexado por $\bar{i}^+$
  Atualizar σ e Ω
  Criar regra R_i
  Atualizar Φ_l e ϱ_i

D. Seleção Incremental de Atributos

O método baseado nos valores médios dos atributos é utilizado para selecionar, de forma incremental, o conjunto mais relevante de atributos. Esses atributos são identificados com base na menor razão entre as médias de seus valores em cada classe [4].

Para isso, inicialmente, os valores médios dos m atributos, $\bar{M} = \{\bar{M}_1, ..., \bar{M}_j, ..., \bar{M}_m\}$, são computados com base nos grupos associados à mesma classe. Para a classe l, a média é obtida por:

$$\bar{M}_j^l = \sum_{i=1}^{c} \frac{\mu_j^i}{s^i} \mid \hat{y}^i \rightarrow l, \quad j = 1, ..., m. \tag{14}$$

em que $\bar{M}_j^l$ é o valor médio do j-ésimo atributo, calculado com base em todas as amostras do grupo pertencentes a classe l;

i é o índice do grupo, c é o número de grupos ativos, μ^i é o centro do grupo e s^i é o número de amostra associadas ao grupo i. A razão entre as médias $\bar{M}$ é computada por

$$\chi_j = \sum_{\forall \hat{y}^i \to l} min\left(\frac{\bar{M}_j^l}{\bar{M}_j^{l-1}}, \frac{\bar{M}_j^{l-1}}{\bar{M}_j^l}\right), \; j = 1, ..., m, \qquad (15)$$

em que χ representa a razão entre os menores e maiores valores médios, expressando uma medida de dispersão relativa. As razões χ_j são usadas para ordenar os atributos x_j em ordem crescente, de forma que atributos com maior dispersão entre as classes sejam considerados mais relevantes para o classificador.

Após o ranqueamento dos atributos, o próximo passo é classificar a amostra x^t, excluindo o atributo menos relevante, $x_{j^\diamond}^t$. Se a amostra for classificada corretamente ($\hat{y}_{exc}^t = y^t$), o atributo $x_{j^\diamond}^t$ é movido do conjunto ativo para o conjunto inativo de atributos.

Se a classificação for incorreta, o atributo mais relevante do conjunto inativo ($x_{j^\bullet}^t$) é selecionado e uma nova classificação é realizada, incluindo agora o atributos $x_{j^\bullet}^t$. Se a amostra for classificada corretamente ($\hat{y}_{inc}^t = y^t$), o atributo $x_{j^\bullet}^t$ é movido do conjunto inativo para o conjunto ativo de atributos. No entanto, se a nova classificação continuar incorreta, o conjunto de atributos atual permanece inalterado.

O Algoritmo 6, ilustra o procedimento de seleção incremental de atributos.

Algoritmo 6. Seleção incremental de atributos.

Calcular $\bar{M}_j^l$ - (14)
Calcular χ_j - (15);
Ranquear os atributos em ordem decrescente;
Encontrar j° - o atributo ativo menos relevante
Calcular $\hat{y}_{exc}$, sem o atributo j° - Algoritmo 2
if $\hat{y} = y$ **then**
  Remover o atributo j° do conjunto de ativos
  Incluir o atributo j° no conjunto de inativos
else
  Encontrar $j^\bullet$ - o atributo inativo mais relevante
  Calcular $\hat{y}_{inc}$, com $j^\bullet$ - Algoritmo 2
if $\hat{y} = y$ **then**
  Remover o atributo $j^\bullet$ do conjunto de inativos
  Incluir o atributo $j^\bullet$ no conjunto de ativos

E. Mesclagem de Grupos

A mesclagem de grupos é realizada utilizando o grau de sobreposição entre grupos que representam a mesma classe, ou seja, quando $\hat{y}^i = \hat{y}^{i^*}$. Para isso, calcula-se a distância Euclidiana entre os centros de todos os pares de grupos. Dois grupos serão mesclados se a distância Euclidiana entre seus centros, μ^{i^*} e μ^i, for a menor entre todos os pares e inferior a um limiar de sobreposição predefinido ρ. Mais especificamente, se:

$$||\mu^{i^*} - \mu^i|| \leq \rho, \;\; \hat{y}^i = \hat{y}^{i^*}, \;\; e \;\; i \neq i^*.$$

O centro dos grupos μ^{i^*} é calculado como a média ponderada entre os centros dos dois grupos que foram mesclados:

$$\mu^{i \cup i^*} = \mu^{i^*} - \frac{s^i}{s^{i^*} + s^i}(\mu^{i^*} - \mu^i), \qquad (16)$$

em que s^i e s^{i^*} representam o número de amostras dos grupos i e i^*. Em seguida, a matriz de dispersão $\Psi^{i^* \cup i}$ é determinada conforme:

$$\Psi^{i^* \cup i} = \frac{\Psi^{i^*} + \Psi^i}{2}, \qquad (17)$$

O número de amostras do grupo mesclado, $s^{i^* \cup i}$, é calculado somando o número de amostras dos dois grupos mesclados:

$$s^{i^* \cup i} = s^{i^*} + s^i, \qquad (18)$$

O número total de grupos criados c, é incrementado em um, em seguida o número de grupos ativos σ e o número de grupos da classe do grupo recém-mesclado Φ é decrementado em um. E então, o vetor de grupos ativos Ω recebe o índice do grupo mesclado:

$$\Omega_\sigma \leftarrow c. \qquad (19)$$

Por fim, a regra mesclada $R^{i^* \cup i}$ é criada, os grupos indexados por i e i^* são removidos de Ω, e os índices são devidamente atualizados. O Algoritmo 7 detalha o procedimento para a mesclagem de grupos.

Algoritmo 7. Mesclagem de grupos e regras.

// i^* - índice do grupo atualizado ou criado
for *i = 1 até p* **do**
  if $i \neq i^*$ *e* $||v_{i^*} - v_i|| \leq \rho$ *e* $\hat{y}_i = \hat{y}_{i^*}$ **then**
    Criar o grupo mesclado $v_{i \cup i^*}$ - (16)
    Calcular $\Psi_{i^* \cup i}$ - (17)
    Calcular $a_{i^* \cup i}$ - (18)
    Atualizar c, σ, Φ
    Atualizar Ω - (19)
    Excluir os grupos indexados por i e i^* de Ω
    Criar a regra mesclada;

F. Exclusão de Grupos

O procedimento de exclusão de grupos utiliza os conceitos de idade e população. A idade de um grupo é definida pelo seu tempo de inatividade, sendo reinicializado para zero, $A_i = 0$, sempre que um grupo é atualizado ou criado. Para o procedimento de exclusão de grupo a idade é obtida por:

$$idade_i^t = t - A_i, \qquad (20)$$

em que i representa o índice do grupo, t é o passo atual, e A_i é o instante de tempo em que o grupo *i*-ésimo foi ativado. Para cada nova amostra, o grupo com maior tempo de inatividade será excluído se:

$$idade_i^t > \omega,$$

sendo ω o limiar de exclusão de grupo por idade.

A população de um grupo é determinada pelo número de amostras associadas a ele. Assim, o grupo i será candidato à exclusão se:

$$\frac{s_i^t}{t} < \Lambda.$$

Para excluir um grupo, identifica-se, em cada nova amostra x^t, o grupo com maior tempo de inatividade, indexado por i^-, conforme:

$$i^- = \arg \max_{i=1,2,\ldots,c^t} idade_i^t. \tag{21}$$

Em seguida, verifica-se se o grupo i^- atende simultaneamente aos critérios de idade e população para exclusão. Além disso, é avaliado se existem dois ou mais grupos que representam a mesma classe do grupo i^-. Por fim, o grupo indexado por i^- será excluído se todas as seguintes condições forem satisfeitas:

$$idade_{i^-}^t > \omega \text{ e } \frac{s_{i^-}^t}{t} < \Lambda \text{ e } \Phi_{i^-} > 1$$

em que Φ_{i^-} é o número de grupos que representam a mesma classe do grupo i^-. Após a exclusão de um grupo, os índices devem ser atualizados. O Algoritmo 8 descreve o procedimento para exclusão de grupos.

Algoritmo 8. Exclusão de grupos e regras.

Atualizar $idade$ do grupo atualizado ou criado - Eq. (20);
Encontrar o índice do grupo menos ativo i^- - Eq. (21);
if $age_{i^-} > \omega$ *and* $a_{i^-}^t < \Lambda$ *and* $\Phi_{i^-} > 1$ **then**
    Excluir o grupo indexado por i^-;
    Atualizar o número de grupos ativos σ
    Atualizar o número de grupos da classe Φ
    Atualizar o vetor de grupo ativos Ω;

G. Hiperparâmetros e Algoritmo de Aprendizagem

O algoritmo de aprendizagem do classificador proposto possui seis hiperparâmetros:

- Ψ_{init}: refere-se à matriz de dispersão inicial, configurada como uma matriz identidade ajustada, com valores variando entre 10^{-1} e 10^{-4}.
- λ: corresponde à taxa de aprendizado, geralmente estabelecida como um valor pequeno. Os valores de λ variam entre 10^{-1} e 10^{-5}.
- ρ: é o parâmetro de mesclagem de novos grupos, mais especificamente o limiar da medida de compatibilidade. ρ é um valor que define o grau de sobreposição entre os grupos, e com base nesse valor se deve permitir ou não a mesclagem de dois grupos. A escolha dos valores de ρ considera os seguintes aspectos: (i) um valor de ρ mais próximo de 1 indica que os grupos mesclados estão mais distantes; (ii) um valor de ρ mais próximo de 0 sugere que os grupos estão mais sobrepostos, ou seja, mais próximos. Na prática, recomenda-se $\rho \in [0.1, 0.4]$.
- ω: é o limiar para excluir um grupo baseado na idade, determinando a idade máxima permitida antes da exclusão. Os valores de ω geralmente variam entre 5 e 200.
- Λ: é o limiar para exclusão de um grupo com base no tamanho da população. Normalmente, Λ é definido como uma porcentagem das amostras processadas, com um intervalo sugerido entre 1 e 4%.
- η: é o parâmetro de procrastinação, responsável por definir o número de amostras necessárias para ativar os grupos. η é expresso por valores inteiros positivos; o valor padrão é 1. Na prática, sugere-se $\eta \in [5, 40]$.

O Algoritmo 9 descreve o fluxo de trabalho do classificador proposto, detalhando as etapas de inicialização do modelo, obtenção da saída, seleção de atributos, atualização da estrutura do modelo e dos grupos.

III. Experimentos e Resultados

Experimentos computacionais foram conduzidos para avaliar e comparar o desempenho do classificador proposto com seis classificadores evolutivos alternativos: ALMMo [13], eFCMG [14], eGNN [15], eOGS [16] e IBEM [17]. As avaliações foram realizadas com base na acurácia média obtida em 30 execuções para cada conjunto de dados, sendo que, em cada execução, as amostras foram embaralhadas utilizando sementes variando de 1 a 30. Para análise estatística, aplicou-se o *T-test* pareado sob a abordagem *all versus one* com um nível de significância de 95%, a fim de comparar o classificador proposto com os demais.

Os classificadores foram avaliados e comparados utilizando quatro conjuntos de dados selecionados aleatoriamente do repositório UCI[1], abrangendo tarefas de classificação binária e multiclasse. A Tabela I apresenta as principais características desses conjuntos, como número de amostras, atributos e classes. As amostras de cada atributo foram normalizadas para o intervalo $[0, 1]$ por meio da normalização $min - max$.

Tabela I. Resumo dos conjuntos de dados utilizados para avaliar os classificadores.

Nome	Abr.	# Amostras	# Atributos	# Classes
Glass	GL	214	09	06
Immunotherapy	IM	90	7	02
Wine	WI	178	13	03
Zoo	ZO	101	16	07

[1]https://archive.ics.uci.edu/

Algoritmo 9. Procedimento de aprendizado do classificador.

Entrada: $x^t,\ y^t$
Saída: $\hat{y}^t$
Inicializar: $\lambda,\ \Psi_{init},\ \omega,\ \rho,\ \Lambda,\ \eta$
for t = *1,2,...* **do**
  Ler amostra x^t
  if t = *1* **then**
    Criar primeiro grupo, classe e regra - Alg. 1
  else
    Calcular saída $\hat{y}^t$ - Alg. 2
    Receber saída desejada $\hat{y}^t$
    //Checar Classificação
    if $\hat{y}^t = \hat{y}^t$ **then**
      Atualizar centro ativo - Alg. 3
    else
      if $\neg\exists y^t = \Theta_l^t$ **then**
        Criar grupo, classe e regra - Alg. 1
      else
        Realizar seleção de atributos - Alg. 6;
        if $\neg\exists y^t = \bar{\Theta}_{\bar{l}}^t$ **then**
          Criar grupo inativo Alg. 4
        else
          Atualizar ou ativar grupo inativo - Alg. 5
      Executar procedimento de união - Alg. 7
      Executar procedimento de exclusão - Alg. 8

O melhor conjunto de hiperparâmetros para cada classificador foi obtido por meio de busca em grade (*grid search*), utilizando 50% das amostras. As amostras restantes foram empregadas para avaliar o desempenho dos classificadores. A Tabela II apresenta os valores ótimos dos hiperparâmetros obtidos para os classificadores em cada conjunto de dados.

A Tabela III ilustra a acurácia média, o valor da estatística do *T-test* e seu respectivo *p-value* para os classificadores avaliados. A maior acurácia média está destacada em azul. Os resultados do *T-test* que indicam superioridade estatística do classificador proposto também estão destacados em azul. Aqueles que apontam inferioridade estatística estão marcados em vermelho, enquanto os casos em que o classificador proposto é estatisticamente comparável ao classificador alternativo estão em preto. O classificador proposto obteve desempenho superior aos classificadores alternativos em três dos quatro conjuntos de dados e foi o segundo melhor no outro. Na média geral, o classificador proposto apresentou o melhor desempenho, seguido por eFCMG, ALMMo, eGNN, eOGS e IBEM.

Os resultados do *T-test* indicam que o classificador proposto apresenta superioridade estatística em relação a todos os classificadores alternativos nos conjuntos de dados *Glass* e *Wine*. No conjunto *Immunotherapy*, o classificador proposto é superior ao IBEM, comparável ao eFCMG, eGNN e eOGS, e inferior ao ALMMo. Por fim, no conjunto *Zoo*, o classificador proposto é estatisticamente superior ao ALMMo, eGNN, eOGS e IBEM, sendo comparável ao eFCMG.

Tabela II. Intervalo dos hiperparâmetros para a busca em grade e melhores valores obtidos para os classificadores evolutivos.

Classificador	Passo	GL	IM	WI	ZO
ALMMo					
ω = 20,...,200	20	160	120	20	100
n^* = 0.1,...,1	0.1	0.30	0.60	1.00	0.40
eFCMG					
ω = 25,...,200	−−	25	50	25	25
ψ = 5,...,35	−−	5	5	5	5
ρ = 0.1,...,0.4	0.1	0.1	0.1	0.1	0.1
Ψ = 0.1,...,0.0001	1.0	0.01	0.1	0.01	0.1
Λ = 1%,...,4%	1.0	1	1	1	1
eGNN					
ω = 50,...,500	50	100	100	50	50
ρ = 0.1,...,1	0.1	0.20	0.90	1.00	0.20
Proposto					
ω = 25,...,200	−−	50	50	25	25
η = 5,...,35	−−	5	5	5	5
ρ = 0.1,...,0.4	0.1	0.1	0.1	0.1	0.1
Ψ_α= 0.1,...,0.0001	1.0	0.01	0.0001	0.01	0.1
Λ = 1%,...,4%	1.0	1	1	1	1
eOGS					
υ = 50,...,500	50	50	50	100	50
α = 0.1,...,1	0.1	0.80	0.60	0.80	0.10
IBEM					
ω = 50,...,500	50	100	50	50	50
ρ = 0.1,...,1	0.1	0.20	0.50	0.10	0.10

Tabela III. Acurácia média, *T-test* e *p-value* na avaliação do desempenho dos classificadores.

Dados	Métrica	Proposto	ALMMo	eFCMG	eGNN	eOGS	IBEM
GL	% Acc	47,57	27,15	44,21	36,82	36,61	25,47
	T-test	—	10,99	2,16	8,70	11,31	15,00
	p-value	—	0,00	0,04	0,00	0,00	0,00
IM	% Acc	72,52	77,03	72,00	67,17	72,46	39,71
	T-test	—	-2,31	0,20	1,95	0,03	14,76
	p-value	—	0,02	0,85	0,06	0,98	0,00
WI	% Acc	91,24	32,80	87,49	64,14	37,92	87,43
	T-test	—	63,93	3,44	18,34	55,37	3,54
	p-value	—	0,00	0,00	0,00	0,00	0,00
ZO	% Acc	56,33	36,86	52,80	44,77	31,30	30,00
	T-test	—	9,02	1,73	5,52	15,58	14,77
	p-value	—	0,00	0,09	0,00	0,00	0,00
Média	% Acc	66,91	43,46	64,12	53,22	44,57	45,65

IV. Conclusão

Este trabalho apresenta um novo classificador *fuzzy* evolutivo que incorpora a seleção incremental de atributos ao processo de aprendizado. O classificador proposto combina um algoritmo de agrupamento com aprendizado participativo gaussiano multivariado e uma estratégia de procrastinação para construir sua estrutura em uma única passagem pelos dados, por meio da inclusão, exclusão, mesclagem e atualização de grupos e regras.

O método de seleção incremental de atributos utiliza a razão média dos atributos dentro das classes para identificar e

selecionar os mais relevantes. A estratégia de procrastinação permite a criação de grupos inativos, que não são imediatamente associados a regras. À medida que novas amostras se tornam disponíveis, esses grupos inativos podem ser ativados, e novas regras geradas. O procedimento para exclusão de grupos e regras combina os conceitos de idade e população. A união de grupos e regras é realizada utilizando o grau de sobreposição dos grupos de mesma classe.

Os resultados experimentais do classificador proposto foram comparados com os de classificadores evolutivos alternativos do estado da arte. Os resultados demonstraram que o classificador proposto superou os modelos alternativos em três dos quatro conjuntos de dados avaliados, com uma melhoria no desempenho geral de classificação entre 3,9% e 50% em relação aos demais classificadores.

Uma limitação da abordagem atual está relacionada ao uso da distância de Mahalanobis, que pode apresentar problemas de singularidade no cálculo do inverso da matriz de dispersão quando o cluster possui um número reduzido de amostras. Para mitigar esse problema, propõe-se o uso combinado de duas medidas de distância: a distância Euclidiana para clusters com poucas amostras e a distância de Mahalanobis para os demais. Além disso, a incorporação de medidas baseadas na distância de Hausdorff pode ampliar a capacidade do modelo em lidar com incertezas, especialmente em cenários com fluxos de dados intervalares. Outro aspecto a ser explorado em trabalhos futuros é a divisão dinâmica de clusters, bem como o aprimoramento da estratégia de procrastinação adotada durante a evolução do modelo.

Disponibilidade de Dados e Códigos

Os conjuntos de dados utilizados nos experimentos — Glass, Immunotherapy, Wine e Zoo — são de acesso público e estão disponíveis no repositório UCI Machine Learning Repository (https://archive.ics.uci.edu/). Visando facilitar a reprodutibilidade dos resultados, esses conjuntos de dados também são disponibilizados juntamente com o código-fonte do classificador proposto, desenvolvido em MATLAB, no seguinte link: https://bit.ly/4guET3L.

Agradecimentos

Alisson Marques da Silva agradece à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio concedido por meio do projeto APQ-03224-24 e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 305452/2025-8. Daniel Leite agradece ao Ministry of Culture and Science of the State of North Rhine-Westphalia pelo apoio por meio do projeto SAIL (nº NW21-059D).

Referências

[1] P. Malhotra, Y. Singh, P. Anand, D. Bangotra, P. Singh, and W. Hong, "Internet of things: Evolution, concerns and security challenges," *Sensors*, vol. 21, no. 5, p. 1809, 2021.

[2] M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, M. I. Tariq, E. De-la Hoz-Franco, and E. De-La-Hoz-Valdiris, "Trends and future perspective challenges in big data," in *Advances in Intelligent Data Analysis and Applications*, J.-S. Pan, V. E. Balas, and C.-M. Chen, Eds. Singapore: Springer Singapore, 2022, pp. 309–325.

[3] E. D. Zamani, C. Smyth, S. Gupta, and D. Dennehy, "Artificial intelligence and big data analytics for supply chain resilience: a systematic literature review," *Annals of Operations Research*, vol. 327, no. 2, pp. 605–632, 2023.

[4] E. Affonso, "Classificação e reconhecimento de padrões: novos algoritmos e aplicações," Tese de doutorado, Programa de Pós-Graduação em Modelagem Matemática e Computacional - PPGMMC CEFET-MG, Belo Horizonte, Brasil, 2 2023.

[5] A. Maria, C. Castiello, M. Luis, and C. Mencar, "Explainable fuzzy systems: Paving the way from interpretable fuzzy systems to explainable ai systems," *Studies in Computational Intelligence*, vol. 970, 2021.

[6] A. M. Silva, W. Caminhas, A. Lemos, and F. Gomide, "A fast learning algorithm for evolving neo-fuzzy neuron," *Applied Soft Computing*, vol. 14, pp. 194–209, 2014, evolving Soft Computing Techniques and Applications.

[7] X. Gu and P. Angelov, "Self-organising fuzzy logic classifier," *Inf. Sci.*, vol. 447, pp. 36–51, 2018.

[8] A. M. Silva, W. Caminhas, A. Lemos, and F. Gomide, "Adaptive Input Selection and Evolving Neural Fuzzy Networks Modeling," *International Journal of Computational Intelligence Systems*, no. 8, p. 1, 2015.

[9] A. M. Silva, W. M. Caminhas, A. P. Lemos, and F. Gomide, *Extended Approach for Evolving Neo-Fuzzy Neural with Adaptive Feature Selection in Decision Making and Soft Computing*, 2014, ch. 5, pp. 651–656.

[10] E. Tavares, A. M. Silva, G. F. Moita, and R. T. N. Cardoso, "A straightforward feature selection method based on mean ratio for classifiers," *Intelligent Decision Technologies*, vol. 15, no. 3, pp. 421–432, 2021.

[11] A. Lemos, W. Caminhas, and F. Gomide, "Multivariable gaussian evolving fuzzy modeling system," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 91–104, 2011.

[12] F. Rodrigues, A. Silva, and A. Lemos, "Evolving fuzzy predictor with multivariable gaussian participatory learning and multi-innovations recursive weighted least squares: efmi," *Evolving Systems*, vol. 13, pp. 1–20, 02 2022.

[13] P. Angelov, X. Gu, and J. Príncipe, "Autonomous Learning Multi-Model Systems from Data Streams," *IEEE Trans. Fuzzy Syst.*, no. 99, p. 1, 2017.

[14] S. Rodrigues, A. M. Silva, and P. V. C. Souza, "eFCMG - an evolving fuzzy classifier with participatory learning and multivariable gaussian for data stream," *Evolving Systems*, vol. 16, pp. 1–22, 2025.

[15] D. Leite, P. Costa, and F. Gomide, "Evolving granular neural networks from fuzzy data streams," *Neural Networks*, vol. 38, pp. 1–16, 2 2013.

[16] D. Leite, G. Andonovski, I. Škrjanc, and F. Gomide, "Optimal rule-based granular systems from data streams," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 3, pp. 583–596, 2020.

[17] D. Leite, P. Costa, and F. Gomide, *Interval Approach for Evolving Granular System Modeling*. New York, NY: Springer New York, 2012, pp. 271–300.