Autonomia e Otimização Multiobjetivo de Sistemas Granulares Fuzzy Evolutivos: Uma Aplicação em Análise EEG

Danyellen Gonçalves* Daniel Leite*

* Departamento de Automação, Universidade Federal de Lavras,
Lavras, MG, Brasil (E-mail: danyellen.g@hotmail.com;
daniel.leite@ufla.br).

Abstract: We present a recursive machine learning method for the optimal construction of rule-based granular models from numerical data streams. We consider a multiobjective function, the specificity of information, model compactness, and data coverage. We use α -level sets over Gaussian membership functions to control the model granularity and operate with adaptive hyper-rectangular granules in nonstationary environment. Rule-based models are developed systematically and can be used for prediction, function approximation, and classification. Numerical estimates are provided by linear piecewise functions from an input-output granular map. An application example in eye-state identification from electroencephalogram (EEG) data shows that the method outperforms other evolving methods.

Resumo: Apresentamos um método recursivo de aprendizado de máquina para a construção ótima de modelos granulares baseados em regras a partir de fluxos de dados numéricos. Consideramos uma função multiobjetivo, a especificidade da informação, a compactação do modelo, e a cobertura dos dados. Usamos conjuntos nível α sobre funções de pertinência Gaussianas para controlar a granularidade do modelo e operar com granulos hiper-retangulares adaptativos em ambiente não-estacionário. O modelo de regras é desenvolvido de forma sistemática e pode ser usado para previsão, aproximação de função, e classificação. Estimativas numéricas são fornecidas por funções lineares por partes a partir de um mapa de grânulos nos espaços de entrada e de saída. Um exemplo de aplicação em identificação do estado dos olhos a partir de dados de eletroencefalograma (EEG) mostra que o método apresentado tem maior acurácia que métodos evolutivos alternativos.

Keywords: Evolving systems; granular computing; online data stream; adaptive systems; electroencephalogram.

Palavras-chaves: Sistemas evolutivos; computação granular; fluxo de dados *on-line*; sistemas adaptativos; eletroencefalograma.

1. INTRODUÇÃO

Desde a última década, as questões de processamento de fluxo de dados e adaptação de modelos em tempo real têm se tornado questões fundamentais. Em termos gerais, o foco da análise de dados mudou do processamento de dados *off-line* para o tratamento recursivo de fluxos de dados *on-line* (Nguyen et al., 2015) (Leite, 2012).

Métodos de Computação Granular têm sido propostos para a obtenção de conhecimento a partir de grandes conjuntos de dados, em princípio, sem valor aparente. Esses métodos examinam o fluxo da informação em ambiente dinâmico e mantém atualizado um modelo (classificador, preditor, controlador) granular, que pode ser compreendido linguisticamente (Dovžan et al., 2015) (Tomazič et al., 2019) (Maciel et al., 2016) (Leite et al., 2013) (Domingos et al., 2016) (Zdešar et al., 2014) (Leite et al., 2015). Os conceitos de grânulo de informação e mapa granular são usados no processo de aprendizado a partir de fluxos de dados. Um grânulo incorpora objetos ou elementos de

um domínio devido a sua similaridade ou funcionalidade (Yao et al., 2013) (Yao, 2013). Um mapa granular é uma associação entre grânulos expressos nos espaços de entrada e de saída. Mapas granulares são encontrados em sistemas baseados em regras (Zadeh, 1997) (Bargiela and Pedrycz, 2005) (Leite et al., 2019).

O conceito de grânulo está relacionado com o conceito de especificidade da informação (Yager, 1992). Um aumento na especificidade da informação tende a aumentar sua utilidade para auxiliar decisões e ações. No entanto, sendo muito específico, corremos o risco de estarmos incorretos, uma vez que a evidência experimental pode testemunhar contra. Além disso, retornamos ao problema original relacionado ao grande volume de dados. Por outro lado, sendo pouco específico, podemos assegurar estarmos corretos. Em teoria da informação, isso é chamado de compromisso especificidade-certeza (Yager, 2008). A granularidade (Pedrycz and Homenda, 2013) de modelos granulares deve representar os dados no sentido de que os grânulos devem maximizar a cobertura dos dados e, ao mesmo tempo,

devem ser razoavelmente específicos, para que se tenha uma descrição significativa (Leite et al., 2019).

Modelos granulares construídos a partir de fluxos de dados podem se fundamentar em áreas como Análise Intervalar, Estatística, Sistemas Fuzzy, Análise de Clusters, Sistemas de Vizinhança. Restrições generalizadas (Zadeh, 2006) são usadas para delimitar e representar grânulos nos diferentes ambientes computacionais. A computação com grânulos permite escolhas dos objetos representativos e das ferramentas de manipulação dos objetos e dos dados. Independente do ambiente, a granulação *on-line* visa manter a essência dos dados como objetos granulares (Leite et al., 2019). Note que a aplicação direta de métodos de aprendizado de máquina à fluxos de dados é, muitas vezes, inviável, pois é difícil ou impossível manter todos os dados na memória. A computação com grânulos adaptativos visa desenvolver gradualmente representações mais abstratas dos dados, e, portanto, mais centradas na forma humana de representação.

O uso de modelos evolutivos tem se tornado comum em modelagem de sistemas (Škrjanc et al., 2019) (Lughofer, 2011) (Soares et al., 2018). Devido a sua capacidade de adaptação paramétrica e estrutural em tempo real, evita-se o reprojetado do modelo quando o ambiente muda. Métodos heurísticos têm sido propostos para desenvolvimento e adaptação incremental de modelos baseados em regras a partir de fluxos de dados numéricos (Zdešar et al., 2014) (Dovzan and Škrjanc, 2011) (Lughofer et al., 2018) e granulares (Leite et al., 2013) (Leite et al., 2012a) (Leite et al., 2019). Soluções persuasivas para objetivos imediatos têm sido alcançadas. No entanto, garantias de que certas condições serão atendidas de maneira ótima ainda são escassas.

Este artigo revisita um método para construção ótima de modelo granular baseado em regras a partir de fluxos de dados numéricos. O sistema granular evolutivo ótimo (eOGS) (Leite et al., 2019) utiliza funções afins por partes e funções de inclusão associadas a grânulos Gaussianos. eOGS fornece estimativas granulares e numéricas de funções variantes no tempo. O modelo eOGS auto-adapta sua estrutura quando um conceito muda ou um novo conceito surge, ao mesmo tempo que minimiza uma função multi-objetivo. Enfatizamos uma aplicação em identificação do estado dos olhos a partir de dados EEG.

2. CONCEITOS ESSENCIAIS

2.1 Granularidade justificável

A análise de grânulos de informação em conjuntos de dados tem sido discutida (Pedrycz and Homenda, 2013) (Pedrycz et al., 2013) (Liu et al., 2018). O princípio da granularidade justificável (Wang et al., 2016) (Ju et al., 2019) é um conceito amplo que visa guiar a descoberta de grânulos a partir de evidências experimentais. Quanto mais dados forem incluídos dentro dos limites de um grânulo, melhor. A cobertura dos dados é uma característica desejada. Queremos que os grânulos incluam todos ou a maioria dos dados. No entanto, os grânulos devem ser tão específicos quanto possível para que tenham significado e utilidade em tomada de decisões. Se muitos dados forem incluídos em um grânulo, o grânulo pode se tornar muito grande. Ao

contrário, se o grânulo for muito pequeno, poucos dados são cobertos. Os requisitos de evidência experimental e especificidade são conflitantes. O método eOGS recorre à otimização multiobjetivo para equilibrar requisitos.

A modelagem de fluxos de dados requer um algoritmo de aprendizado incremental que mantém um resumo dos dados. A ideia é capturar informações espaciais e temporais em dados numéricos $x^{[h]} = (x_1, \dots, x_j, \dots, x_n)^{[h]}$, $h = 1, \dots$ usando um conjunto de grânulos $\gamma = \{\gamma^1, \dots, \gamma^c\}$. Um grânulo γ^i é escolhido para acomodar uma amostra, isto é, para incluir $x^{[h]}$. Se $x^{[h]}$ não está suficientemente relacionado a nenhum γ^i , então um novo grânulo γ^{c+1} é criado - expandindo a coleção atual γ . Portanto, γ é um modelo granular que descreve o fluxo de dados $x^{[h]}$, $h = 1, \dots$

Grânulos de natureza intervalar dependem essencialmente dos limites inferior, l_j^i , e superior, L_j^i , de hiper-retângulos alinhados aos eixos. Os extremos de grânulos são determinados a partir de conjuntos nível- α . Suponha uma função Gaussiana $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, que representa a j -ésima característica do i -ésimo grânulo. O intervalo de valores em que as amostras de dados estão, na maioria dos casos, é dado pelo corte α de Π_j^i , isto é, por $[l_j^i, L_j^i]$. A extensão cilíndrica do conjunto nível- α no espaço do produto Cartesiano forma um grânulo. Grânulos intervalares cobrem os dados, e funções de pertinência Gaussianas mantêm a essência da informação do fluxo de dados.

Dependendo da variância de Π_j^i e do corte α , grânulos e modelos baseados em regras podem alcançar qualquer especificidade. Por exemplo, se um algoritmo de aprendizado é usado para adaptar recursivamente μ_j^i e σ_j^i , então Π_j^i é flexível para rastrear não-estacionariedades em fluxos de dados. Ao mesmo tempo, podemos usar o valor do corte α para definir o tamanho dos intervalos $[l_j^i, L_j^i]$, ou seja, para controlar a especificidade dos grânulos.

2.2 Especificidade da informação

Especificidade refere-se à quantidade de informação de um grânulo (Yager, 2008). Uma medida de especificidade se aproxima do seu valor máximo à medida que um objeto representa um único elemento. Valores de especificidade variam em $[0, 1]$; 0 significa que todos os pontos de um domínio são igualmente possíveis, e 1 implica singularidade. Seja Π um subconjunto de X . Yager (Yager, 2008) define especificidade sobre domínios contínuos como

$$\text{Sp}(\Pi) = \int_0^{\alpha_{max}} F(\lambda(\Pi_\alpha)) d\alpha, \quad (1)$$

onde $\Pi_\alpha = \{x : \Pi(x) \geq \alpha\}$ é o conjunto nível- α de Π ; α_{max} é a altura de Π ; λ é uma medida monotônica; e $F : [0, 1] \rightarrow [0, 1]$ é uma função tal que $F(0) = 1$; $F(1) = 0$; e $0 \leq F(x_1) \leq F(x_2)$ para $x_1 > x_2$.

Seja Π uma função Gaussiana em X com valor modal μ e dispersão σ ,

$$\Pi = e^{-(x-\mu)^2/2\sigma^2}. \quad (2)$$

Π é sempre maior que 0, e $\int_{-\infty}^{\infty} \Pi dx$ não é necessariamente 1 já que Π é normal. Π_α é o intervalo nível- α de Π . Manipulando (2), $\alpha > 0$ corta Π em $\mu \pm \sqrt{-2\sigma^2 \ln(\alpha)}$.

Seja $X = [c, d]$ um domínio limitado. Usando a definição (1), e Π_α , obtemos a especificidade de Π em função de σ e α conforme

$$\text{Sp}(\Pi) = 1 - \frac{2\sqrt{2}\sigma\sqrt{-\ln(\alpha)}}{d - c}. \quad (3)$$

Seja $\Pi^i = \{\Pi_1^i, \dots, \Pi_j^i, \dots, \Pi_n^i\}$ um conjunto de funções Gaussianas, e $X^n = X_1 \times \dots \times X_j \times \dots \times X_n$. Através de cortes α em cada dimensão, obtemos um grânulo γ^i em X^n . Definimos a especificidade de um grânulo γ^i n -dimensional como a média da medida de especificidade de cada um de seus n componentes,

$$\text{Spa}(\gamma^i) = 1 - \frac{2\sqrt{2}}{n} \sum_{j=1}^n \frac{\sigma_j^i \sqrt{-\ln(\alpha)}}{d_j - c_j}. \quad (4)$$

2.3 Estratégia de decisão

Os critérios usados para guiar a adaptação dos meta-parâmetros $\mathbb{P}$ do modelo eOGS são: erro numérico (E_n), erro granular (E_g), especificidade do mapeamento granular ($\text{Spa}(\gamma)$), e número total de regras (c), ver (Leite et al., 2019). Os meta-parâmetros $\mathbb{P}$ serão apresentados oportunamente. Geralmente, queremos reduzir o erro, numérico e granular, aumentar a especificidade, e manter a base de regras compacta. Porém, um número menor de regras tende a produzir mapas granulares menos específicos, e pode gerar estimativas melhores ou piores, dependendo dos dados.

A função multiobjetivo a ser minimizada é

$$F(\mathbb{P}) = \min [f_1(\mathbb{P}), f_2(\mathbb{P}), f_3(\mathbb{P}), f_4(\mathbb{P})] \\ \text{sujeito a } \mathbb{P} \in \Omega \quad (5)$$

onde Ω é o espaço de parâmetros e

$$E_n \triangleq f_1(\mathbb{P}), \\ E_g \triangleq f_2(\mathbb{P}), \\ -\text{Spa}(\gamma) = -\sum_{i=1}^c \text{Spa}(\gamma^i) \triangleq f_3(\mathbb{P}), \\ c \triangleq f_4(\mathbb{P}).$$

Os objetivos em (5) competem; logo, a solução não é única. A melhoria de um objetivo pode degradar os outros. Os parâmetros $\mathbb{P}$ são apresentados na próxima seção.

No método das ϵ -restrições (Chankong and Haimes, 1963), um objetivo prioritário, f_s , é escolhido para ser otimizado, enquanto os outros, f_t , são convertidos em restrições, definindo um limite superior para cada um deles. O problema (5) assume a forma

$$F(\mathbb{P}) = \min f_s(\mathbb{P}) \\ \text{sujeito a } f_t(\mathbb{P}) \leq \epsilon_t, \forall t, t \neq s \\ \mathbb{P} \in \Omega \quad (6)$$

cuja solução é fracamente Pareto ótima (Chankong and Haimes, 1963). Uma vez que um objetivo prioritário e

os limites das restrições são escolhidos, o algoritmo de aprendizado e o modelo eOGS podem ser vistos como uma heurística que resolve (6).

3. MODELO GRANULAR EVOLUTIVO ÓTIMO

Modelos granulares evolutivos fornecem estimativas numéricas e granulares de funções variantes no tempo (Leite et al., 2012a) (Leite et al., 2010). Um algoritmo de aprendizado *on-line* constrói a estrutura do modelo e adapta os parâmetros sempre que novos padrões são encontrados. O algoritmo detecta mudanças, e lida com incerteza.

O princípio de funcionamento consiste em incluir dados semelhantes em grânulos para simplificar os cálculos. O modelo eOGS é interpretável. Um mapa granular de X para Y é uma regra que atribui a cada grânulo de X , um grânulo em Y . O modelo eOGS codifica um mapa granular como regras Se-Então. A saída granular é dada por uma função de inclusão ou pelos limites dos grânulos de saída. A saída numérica é dada por funções algébricas.

3.1 Regras

Um grânulo eOGS, γ^i , é definido pelos limites de conjuntos nível- α associados às funções de pertinência $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, $j = 1, \dots, n$. O grânulo γ^i é descrito pela regra

$$R^i: \text{SE } (l_1^i \leq x_1 \leq L_1^i) \text{ E } \dots \text{ E } (l_n^i \leq x_n \leq L_n^i)$$

$$\text{ENTÃO } (u_1^i \leq y_1 \leq U_1^i) \text{ E } \bar{y}_1^i = p_1^i(x_1, \dots, x_n) \text{ E}$$

...

$$(u_m^i \leq y_m \leq U_m^i) \text{ E } \bar{y}_m^i = p_m^i(x_1, \dots, x_n),$$

onde l_j^i e L_j^i , $j = 1, \dots, n$; $i = 1, \dots, c$, são os j -ésimos limites inferior e superior do atributo x_j da i -ésima regra; u_k^i e U_k^i , $k = 1, \dots, m$, são os k -ésimos limites inferior e superior da saída y_k ; e p_k^i são funções numéricas afins,

$$\bar{y}_k^i = p_k^i(x_1, \dots, x_n) = a_{0k}^i + \sum_{j=1}^n a_{jk}^i x_j. \quad (7)$$

O grânulo γ^i contém funções de pertinência Gaussianas $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, $j = 1, \dots, n$, em sua representação interna - uma informação adicional que resume os dados pertencentes a γ^i . Valores modais, μ_j^i , e dispersões, σ_j^i , são obtidos recursivamente a partir do fluxo de dados. Além disso, o algoritmo Mínimos Quadrados Recursivo (RLS) (Leite et al., 2015) é usado para determinar os coeficientes a_{jk}^i , $j = 0, \dots, n$, $k = 1, \dots, m$, das funções p_k^i sempre que a i -ésima regra está ativa para uma amostra.

3.2 Inicialização de Stiegler

Grânulos e regras são criados se $x^{[h]} \notin [l_j^i, L_j^i] \forall i$ e algum j ; ou $y^{[h]} \notin [u_k^i, U_k^i] \forall i$ e algum k . Os limites dos intervalos dependem do valor de α que corta Π_j^i e Π_k^i . Se $\alpha \rightarrow 0^+$, então $[l_j^i, L_j^i] \forall j$ e $[u_k^i, U_k^i] \forall k$ cobrem todo o espaço dos dados. Pelo contrário, se $\alpha \rightarrow 1^-$, então γ^i é um ponto.

Uma abordagem para iniciar os parâmetros de um grânulo consiste do uso da função Gaussiana padrão de Stigler (Stigler, 1982). O novo grânulo γ^{c+1} tem valor modal

$$\mu_{j,k}^{c+1} = (x_j, y_k)^{[h]} \quad \forall j, k, \quad (8)$$

e dispersão

$$(\sigma_{j,k}^{c+1})^2 = 1/2\pi \quad \forall j, k. \quad (9)$$

Um hiper-retângulo centrado em $(x, y)^{[h]}$ é obtido no espaço do produto Cartesiano a partir de cortes α unidimensionais. Os limites dos grânulos são dados como

$$[l_j^{c+1}, L_j^{c+1}] = x_j^{[h]} \pm \sqrt{-2(\sigma_j^{c+1})^2 \ln(\alpha)} \quad (10)$$

e

$$[u_k^{c+1}, U_k^{c+1}] = y_k^{[h]} \pm \sqrt{-2(\sigma_k^{c+1})^2 \ln(\alpha)}, \quad (11)$$

$0 < \alpha < 1$. Os coeficientes polinomiais são

$$a_{0k}^{c+1} = y_k^{[h]}, \quad \forall k; \text{ and } a_{jk}^{c+1} = 0, \quad \forall k, j = 1, \dots, n. \quad (12)$$

A especificidade dos grânulos é obtida a partir de (4).

A abordagem de inicialização do Stiegler exige algumas iterações para que o novo grânulo encolha, e seja dimensionado conforme os demais. Novas amostras dentro dos limites do novo grânulo são usadas para adaptar seu valor modal, dispersão, e coeficientes consequentes.

3.3 Adaptação

Fluxos de dados mudam com o tempo. A adaptação das regras do eOGS consiste em expandir ou contrair os limites dos grânulos e, simultaneamente, alterar os coeficientes das funções locais.

Considere os dados em uma janela de tempo, $(x, y)^{[h-\nu]}$, $(x, y)^{[h-\nu+1]}$, ..., $(x, y)^{[h]}$; ν é o comprimento da janela, h é o passo de tempo. Se um novo $x^{[h]}$ pertence a $[l_j^i, L_j^i] \quad \forall j$, e $y^{[h]}$ pertence a $[u_k^i, U_k^i] \quad \forall k$ e algum i , então os parâmetros das Gaussianas locais, Π^i , são atualizados a partir de

$$\mu_{j,k}^i(\text{new}) = \frac{\nu^i \mu_{j,k}^i(\text{old}) + (x_j, y_k)^{[h]}}{\nu^i + 1}, \quad \forall j, k \quad (13)$$

e

$$(\sigma_{j,k}^i(\text{new}))^2 = \beta^i (\sigma_{j,k}^i(\text{old}))^2, \quad \forall j, k, \quad (14)$$

onde

$$\beta^i \triangleq \frac{\nu^i (\mu_{j,k}^i - (l_j, u_k)) + \Psi(|\mu_{j,k}^i - (x_j, y_k)^{[h]}|)}{(\nu^i + 1)(\mu_{j,k}^i - (l_j, u_k))},$$

ν^i é o número de amostras pertencentes a γ^i das últimas ν amostras; $\Psi = 2$ é um valor padrão. Valores superiores ou inferiores para o parâmetro Ψ forçam a expansão ou contração das Gaussianas ao longo das iterações.

O conjunto nível α do grânulo atualizado, γ^i , é

$$[(l_j^i, u_k^i), (L_j^i, U_k^i)] = \mu_{j,k}^i(\text{new}) \pm \sqrt{-2 \ln(\alpha) (\sigma_{j,k}^i(\text{new}))^2} \quad (15)$$

$\forall j, k$. Além disso, a especificidade do grânulo, $\text{Spa}(\gamma^i)$, é obtida como em (4). Dado um objetivo prioritário f_s , o método das ϵ -restrições (6) encontra α e Ψ para satisfazer as restrições f_t e minimizar f_s . Os valores de α e Ψ são usados em (14) e (15) para atualizar as dispersões Π_j^i e Π_k^i e determinar os limites $[l_j^i, L_j^i]$ e $[u_k^i, U_k^i]$. Usuários podem manifestar preferência escolhendo o objetivo principal e definindo valores admissíveis para as restrições $\epsilon_t \forall t$.

Observe que os procedimentos de adaptação descritos são recursivos, ou seja, os valores são acumulados. Portanto, uma amostra $(x, y)^{[h]}$ pode ser descartada após o processamento. O algoritmo RLS (Leite et al., 2015) adapta os coeficientes $a_{jk}^i \quad \forall j, k$, para amostras que ativam γ^i . Observe também que o tamanho da janela de tempo, ν , significa a vida útil da informação na memória de curto prazo de modelos eOGS. Os modelos consideram as últimas ν amostras apenas, para evoluir.

3.4 Fusão

A fusão de grânulos, digamos γ^{i_1} e γ^{i_2} , é útil para reduzir o número de regras e eliminar grânulos parcialmente sobrepostos. Considere o argumento mínimo da norma 2 da diferença entre pontos médios,

$$\arg \min_{i_1, i_2=1, \dots, c; i_1 \neq i_2} \frac{\|\mu^{i_1} - \mu^{i_2}\|}{n}, \quad (16)$$

onde n é o número de atributos, ou dimensões de μ . Obtidos os dois grânulos mais próximos, γ^{i_1} e γ^{i_2} , se

$$\frac{\|\mu^{i_1} - \mu^{i_2}\|}{n} \leq \omega, \quad (17)$$

sendo ω um limiar variante no tempo, então os grânulos são mesclados. Uma heurística de adaptação para ω é descrita na seção Metodologia.

O grânulo resultante, γ^{c+1} , é formado por $n + m$ funções de pertinência Gaussianas, cujos valores modais dependem da quantidade de amostras que cada grânulo representa. Formalmente,

$$\mu_{j,k}^{c+1} = \frac{\nu^{i_1} \mu_{j,k}^{i_1} + \nu^{i_2} \mu_{j,k}^{i_2}}{\nu^{i_1} + \nu^{i_2}}, \quad \forall j, k. \quad (18)$$

A abordagem máxima é usada para a dispersão,

$$\sigma_{j,k}^{c+1} = \max(\sigma_{j,k}^{i_1}, \sigma_{j,k}^{i_2}), \quad \forall j, k. \quad (19)$$

A ideia é preservar informações das amostras na região de fusão. Dados (18) e (19), os limites e especificidade dos grânulos são obtidos de (15) e (4), por analogia. Coeficientes são definidos como

$$a_{jk}^{c+1} = \frac{a_{jk}^{i_1} + a_{jk}^{i_2}}{2}, \quad \forall j, k. \quad (20)$$

3.5 Exclusão

Regras inativas por várias iterações podem ser excluídas. Isso pode significar que o sistema mudou, e a remoção de grânulos é interessante para manter a base de regras compacta. Relembre que ν^i é a quantidade de amostras que ativaram γ^i , considerando a janela ν , $x^{[h-\nu]}, \dots, x^{[h]}$. Para grânulos inativos, $\nu^i = 0$, e o respectivo grânulo e regra são removidos.

3.6 Função intervalar

A imagem de γ^i através da função real p_k^i é

$$p_k^i([l_1^i, L_1^i], \dots, [l_n^i, L_n^i]) = \{p_k^i(x_1, \dots, x_n) : x_j \in [l_j^i, L_j^i], j = 1, \dots, n\}.$$

Geralmente, a imagem de γ^i através de p_k^i não é um hiper-retângulo, e pode ser difícil de ser obtida na forma fechada. Na prática, p_k^i pode ser aproximado por uma função de inclusão, P_k^i , que é um hiper-retângulo na faixa de p_k^i , ou seja $[u_k^i, U_k^i]$.

Uma função intervalar P_k^i é chamada de função de inclusão se $p_k^i \subseteq P_k^i \forall i$. Funções de inclusão não são exclusivas; elas dependem de como escolhemos P (Jaulin et al., 2001). Uma função de inclusão P_k^i é ótima se ela for a casca convexa de p_k^i . Em outras palavras, a função de inclusão ótima para p_k^i é o menor hiper-retângulo P_k^{i*} que contém p_k^i . P_k^{i*} é único. Sua especificidade é a mais alta possível que garante inclusão.

Assuma p_k^i aumentando monotonicamente em $[l_j^i, L_j^i]$, $j = 1, \dots, n$. Então podemos obter p_k^i de

$$p_k^i(x) = [p_k^i(l_1^i, \dots, l_n^i), p_k^i(L_1^i, \dots, L_n^i)].$$

Consequentemente, para qualquer $x \in \gamma^i$, $p_k^i(x) \subseteq [p_k(l^i), p_k(L^i)]$. Para funções monotônicas decrescentes,

$$p_k^i(x) = [p_k^i(L_1^i, \dots, L_n^i), p_k^i(l_1^i, \dots, l_n^i)].$$

Adotamos

$$p_k^i(x \in \gamma^i) = a_{0k}^i + \sum_{j=1}^n a_{jk}^i [l_j^i, L_j^i], \quad (21)$$

onde $(a_{0k}^i, \dots, a_{nk}^i)$ são intervalos degenerados. Se a inclinação $a_{jk}^i < 0$, então os limites devem ser invertidos, ou seja, $[L_j^i, l_j^i]$.

3.7 Algoritmo de aprendizado

O algoritmo de aprendizado de modelos eOGS é resumido da seguinte forma:

INÍCIO

Definir $\mathbb{P} = \{\alpha, \Psi, \nu, \omega\}$ (valores padrão ou escolhidos);
Definir um objetivo prioritário f_s ;
Definir valores admissíveis para as restrições $\epsilon_t \forall t$;
for $h = 1, \dots$

```
Ler dados de entrada  $x^{[h]}$ ;  
if  $h = 1$   
  Criar grânulo  $\gamma^1$  e regra  $R^1$  (Eqs. (8)-(12) e (4));  
  Fornecer previsão numérica  $\bar{y}^{[h]}$  e granular  $[u^1, U^1]^{[h]}$ ;  
else  
  Fornecer previsão numérica  $\bar{y}^{[h]}$  e granular  $[u^i, U^i]^{[h]}$ ;  
  // A saída atual  $y^{[h]}$  se torna disponível;  
  if  $x_j^{[h]} \notin [l_j^i, L_j^i] \forall i$  e algum  $j$   
    Criar grânulo  $\gamma^{c+1}$  e regra  $R^{c+1}$  (Eqs. (8)-(12) e (4));  
  else  
    Adaptar grânulos ativos  $\gamma^i$  (Eqs. (13)-(15) e (4));  
  end  
end  
Excluir grânulos e regras inativos;  
Mesclar grânulos e regras (Eqs. (16)-(20), (4), (15));  
Adaptar  $\mathbb{P} = \{\alpha, \Psi, \nu, \omega\}$  para minimizar  $f_s$  respeitando  $\epsilon_t \forall t$ ;  
end  
FIM
```

4. METODOLOGIA

São apresentadas medidas de erro para avaliar a acurácia do modelo. Um procedimento automático para adaptar os meta-parâmetros de eOGS é apresentado.

4.1 Índices de desempenho

O erro quadrático médio das previsões numéricas,

$$RMSE = \frac{1}{H} \sum_{h=1}^H \sqrt{\frac{1}{m} \sum_{k=1}^m (y_k^{[h]} - \bar{y}_k^{[h]})^2}, \quad (22)$$

em que H é o número atual de iterações e m é o número de saídas; é uma medida de acurácia para comparar modelos diferentes para um conjunto de dados específico.

Uma medida de erro granular (Leite et al., 2019) que leva em conta a inclusão dos dados e o invólucro gerado é

$$MGE = \frac{1}{kH} \sum_{h=1}^H \sum_{k=1}^m 1 - \mathfrak{z}_k^{[h]} (1 - (U_k^{[h]} - u_k^{[h]})), \quad (23)$$

em que

$$\mathfrak{z}_k^{[h]} = \begin{cases} 1 & \text{se } y_k^{[h]} \in [u_k, U_k]^{[h]} \\ 0 & \text{caso contrário.} \end{cases} \quad (24)$$

Se a saída real $y_k^{[h]}$ estiver fora dos limites de previsão $[u_k, U_k]^{[h]}$, então $\mathfrak{z}_k^{[h]} = 0$, produzindo o MGE máximo. O índice MGE é menor que 1 somente se a previsão granular incluir $y_k^{[h]}$. Quanto mais estreitos forem os limites $[u_k, U_k]^{[h]}$ que inclui $y_k^{[h]}$, menor será o MGE .

4.2 Operação Automática

Uma forma típica do problema de otimização (6) assume o erro numérico como o objetivo primário, ou seja,

$$\begin{aligned} F(\mathbb{P}) &= \min RMSE \\ \text{sujeito a } MGE &\leq \epsilon_1 \\ c &\leq \epsilon_2 \\ Spa(\gamma) &\geq \epsilon_3 \end{aligned} \quad (25)$$

O $RMSE$ pode ser trocado com qualquer das restrições. Se uma restrição for violada, o algoritmo deve responder o mais rápido possível por meio dos parâmetros $\mathbb{P} = \{\alpha, \Psi, \nu, \omega\}$. Os valores padrão para $\mathbb{P}$ são $\alpha = 0, 1$, $\Psi = 2$, $\nu = 500$ e $\omega = 0, 01$. Se um certo conjunto de amostras estiver disponível, um fluxo de dados pode ser simulado em questão de segundos e, talvez, um conjunto mais apropriado $\mathbb{P}$ possa ser encontrado.

Durante a operação *on-line* do algoritmo eOGS, se uma restrição é violada, os parâmetros são alterados conforme:

- Se $MGE > \epsilon_1$, então α é aumentado, e σ^2 é diminuído para novos grânulos, mas mantido dentro de uma faixa mínima e de Stiegler, isto é, $0, 01 \leq \sigma^2 \leq 1/2\pi$.
- Se $c > \epsilon_2$, então α é diminuído, e ω é aumentado. Após algumas iterações, quando c assume um valor viável, ω é redefinido para o padrão.
- Se $Spa(\gamma) < \epsilon_3$, então α é aumentado, e Ψ é diminuído. Após algumas iterações, quando $Spa(\gamma)$ assume um valor viável, Ψ é redefinido para o padrão.

Além disso,

- Se $RMSE > \epsilon_4$, então α é diminuído, e σ^2 é aumentado para novos grânulos, mas mantido dentro da faixa $0, 01 \leq \sigma^2 \leq 1/2\pi$.

Se os valores ϵ_i forem definidos de forma irreal, os parâmetros $\mathbb{P}$ alcançam um limite e a melhor solução possível é apresentada (Leite et al., 2019).

5. EXEMPLO DE APLICAÇÃO

5.1 Identificação do Estado dos Olhos

Atualmente existe grande interesse na área de classificação de dados de eletroencefalografia (EEG). Há uma variedade enorme de aplicações possíveis a partir da detecção de padrões de pensamento, tais como controle de cadeira de rodas (Carlson and del R Millan, 2013), sistemas de saúde e biomédicos (Ossmy et al., 2011), entre outros. Este trabalho foca em detecção do estado dos olhos a partir das leituras de canais EEG. A base de dados intitulada Identificação do Estado dos Olhos, disponível no UCI Machine Learning Repository, foi considerada (Rösler and Suendermann, 2013).

Os sinais EEG foram obtidos a partir de medições contínuas, por 117 segundos, utilizando o dispositivo Neuroheadset Emotiv EEG, com 14 eletrodos, localizado nas posições F7, F8, F3, F4, FC5, FC6, T7, T8, P3, P4, P7, P8, O1 e O2 de acordo com o Sistema Internacional 10-20 (Malmivuo and Plonsey, 1995). O estado dos olhos (valor de saída 'desejado' de modelos classificadores) foi detectado por meio de uma câmera, ao longo das medições, e adicionados manualmente à base de dados após análises de vídeo. O estado '1' indica olho fechado, e o estado '0' se refere ao olho aberto.

Nesta base de dados existem quatro amostras (899, 10387, 11510 e 13180) contendo erros notáveis. Seus valores são muito superiores àqueles possíveis, i.e., valores dentro da faixa admissível dos canais EEG. Estes dados foram considerados *outliers* e a amostra completa foi excluída antes dos experimentos para construção de modelos classificadores. Portanto, apenas 14976 amostras são utilizadas no experimento, sendo 6722 relativas ao estado fechado do olho, e 8254 ao estado aberto.

As amostras são apresentadas sequencialmente ao algoritmo eOGS e à algoritmos alternativos, usados para comparação de desempenho. São algoritmos alternativos: *interval based evolving modeling* (IBeM) (Leite et al., 2012b), *evolving granular neural network* (eGNN) (Leite et al., 2013) e *fuzzy set based evolving modeling* (FBeM) (Leite et al., 2012a). Os modelos são construídos sem que haja regras, grânulos ou clusters iniciais, nem pré-treinamento. Os meta-parâmetros iniciais de eOGS são adotados conforme o padrão, exceto $\nu = 1000$. Isso significa que queremos manter uma certa memória de informações passadas ao invés de excluir regras inativas. Os parâmetros $\rho = 0, 25$, $h_r = 300$, $\eta = 2$ foram utilizados nos três algoritmos alternativos: IBeM, eGNN e FBeM.

5.2 Resultados

Seja o erro numérico de estimação o objetivo principal. O problema de otimização é

$$\begin{aligned} F(\mathbb{P}) &= \min RMSE \\ \text{sujeito a } MGE &\leq 1 \\ c &\leq 8 \\ Spa(\gamma) &\geq 0, 6 \end{aligned} \quad (26)$$

A Fig. 1 mostra a aproximação numérica do estado dos olhos; a evolução do nível α , do número de regras, e da especificidade ao longo das iterações; e o índice de erro $RMSE$. Em particular, não estamos interessados na aproximação granular nesta aplicação, pois se trata de um problema de classificação binário. Logo, em (26) escolhemos $MGE \leq 1$. A Fig. 1 também mostra as Gaussianas da saída. O segundo gráfico expande a aproximação do primeiro gráfico dentro da faixa [2000, 3500] para um maior detalhamento.

Nota-se na Fig. 1 que o índice $RMSE$ reduziu para 0,0690 durante as iterações, ou seja, o modelo está efetivamente aprendendo em tempo real. Além disso, o número de regras c , e $Spa(\gamma)$ praticamente não violam suas restrições. Quando isso acontece, o algoritmo de aprendizado reage imediatamente, e retoma a factibilidade. Os dados EEG foram processados no ambiente MATLAB. O tempo total da CPU para processar 14976 amostras de dados usando um processador i5-7200U de 2,5-2,7GHz com 8GB de RAM foi de 18,1 segundos, o que indica que o algoritmo pode operar em tempo real considerando fluxos de dados de alta frequência.

A estimativa numérica da classe, mostrada nos dois primeiros gráficos da Fig. 1, fornece a indicação do estado dos olhos, i.e., '0' ou '1'. Valores acima de 0,5 indicam olhos fechados, e valores abaixo de 0,5 se referem a olhos abertos.

Figura 1. Função de aproximação numérica eOGS, evolução de α , número de regras, especificidade, índice de erro, e Gaussianas de saída

eOGS é capaz de separar as amostras em classes a partir de um posicionamento particular da fronteira de decisão que maximiza a separação das classes. Note que os dados não se aproximam muito do ponto central, 0,5, e que há um equilíbrio na aproximação por ambos os lados. Erros de classificação são momentâneos, durante transições, visto a imprevisibilidade do momento do piscar de olhos e um atraso de comunicação e processamento mínimo.

No último gráfico da Fig. 1, nota-se um retrato das Gaussianas de saída na última iteração. Neste instante, 4 regras são suficientes para classificar os dados EEG. O número médio de regras ao longo da evolução foi de $c = 4,5$. As Gaussianas das extremidades mapeiam dados comuns dos eletrodos e, portanto, têm suas dispersões reduzidas. As duas Gaussianas intermediárias são mais flexíveis ao fluxo de dados, e arrastadas ao longo do tempo no sentido de maximizar a fronteira de decisão.

A Tabela 1 compara o método eOGS e métodos evolutivos alternativos em termos do número de regras, RMSE e acurácia de classificação. eOGS classificou dados EEG com acurácia de 99,73%, e superou os demais métodos.

Tabela 1. Avaliação de diferentes métodos evolutivos na identificação do estado dos olhos

Modelo	# Regras	RMSE	Acurácia
IBeM	4,8	0,2673	91,76%
eGNN	7,1	0,2074	95,70%
FBeM	2,7	0,1062	98,86%
eOGS	4,5	0,0690	99,73%

6. CONCLUSÃO

Este estudo apresentou um algoritmo de aprendizado recursivo para a construção ótima de modelos granulares baseados em regras a partir de fluxos de dados. Sua formulação multiobjetivo considera a especificidade da informação, a compactação do modelo, e estimativas numéricas e granulares. Conjuntos nível- α obtidos de funções de pertinência são usados para definir a granularidade do modelo em ambiente não-estacionário.

Modelos são construídos de forma sistemática. Estimativas numéricas são fornecidas por funções lineares por partes. Estimativas granulares podem ser produzidas por função de inclusão relacionada a mapas granulares. Um conjunto de dados EEG referentes ao estado dos olhos de uma pessoa foi utilizado na comparação do algoritmo eOGS com outros algoritmos evolutivos (FBeM, IBeM e eGNN). A definição de meta-parâmetros foi feita de maneira autônoma. Os resultados indicam que o algoritmo eOGS é aplicável à classificação de séries temporais de eletroencefalograma e supera outros métodos evolutivos em acurácia.

Em trabalhos futuros será analisado o funcionamento de eOGS na classificação de dados EEG provenientes da execução de outras atividades. Serão considerados problemas multi-classe e controle de dispositivos físicos a partir da associação de ações de controle aos rótulos de grânulos.

AGRADECIMENTOS

Este trabalho recebeu apoio do Instituto Serrapilheira, número do processo Serra – 1812-26777. A primeira autora agradece à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pela bolsa de mestrado. O segundo autor agradece à FAPEMIG, processo APQ-03384-18.

REFERÊNCIAS

Bargiela, A. and Pedrycz, W. (2005). Granular mappings. *IEEE Trans Syst Man Cybern - Part A*, 35(2), 292–297.

- Carlson, T. and del R Millan, J. (2013). Brain-controlled wheelchairs: a robotic architecture. *IEEE Robot Autom Mag*, 20(1), 65–73.
- Chankong, V. and Haimes, Y. (1963). *Multiobjective Decision Making Theory and Methodology*. North-Holland, NY.
- Domingos, D., Camargo, G., and Gomide, F. (2016). Autonomous fuzzy control and navigation of quadcopters. *IFAC - PapersOnLine*, 49(5), 73–78.
- Dovžan, D., Logar, V., and Škrjanc, I. (2015). Implementation of an evolving fuzzy model (efumo) in a monitoring system for a waste-water treatment process. *IEEE Trans Fuzzy Syst*, 23(5), 1761–1776.
- Dovzan, D. and Skrjanc, I. (2011). Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes. *ISA Trans*, 50(2), 159–169.
- Jaulin, L., Keiffer, M., Didrit, O., and Walter, E. (2001). *Applied Interval Analysis*. Springer-Verlag, London.
- Ju, H., Pedrycz, W., Li, H., Ding, W., Yang, X., and Zhou, X. (2019). Sequential three-way classifier with justifiable granularity. *Knowl.-Based Syst*, 163, 103–119.
- Leite, D. (2012). Evolving granular systems. *PhD thesis, University of Campinas, School of Electrical and Computer Engineering*.
- Leite, D., Ballini, R., Costa, P., and Gomide, F. (2012a). Evolving fuzzy granular modeling from nonstationary fuzzy data streams. *Evolving Systems*, 3(2), 65–79.
- Leite, D., Costa, P., and Gomide, F. (2010). Granular approach for evolving system modeling. In *Lecture Notes in Artificial Intelligence*, 340–349. Springer, Heidelberg.
- Leite, D., Costa, P., and Gomide, F. (2012b). Interval approach for evolving granular system modeling. In *Learning in Non-Stationary Environments*, 271–300. Springer, New York.
- Leite, D., Costa, P., and Gomide, F. (2013). Evolving granular neural networks from fuzzy data streams. *Neural Networks*, 38, 1–16.
- Leite, D., Palhares, R., Campos, V., and Gomide, F. (2015). Evolving granular fuzzy model-based control of nonlinear dynamic systems. *IEEE Trans Fuzzy Syst*, 23(4), 923–938.
- Leite, D., Andonovski, G., Skrjanc, I., and Gomide, F. (2019). Optimal rule-based granular systems from data streams. *IEEE Trans on Fuzzy Syst*, 1–14, DOI: 10.1109/TFUZZ.2019.2911493.
- Liu, S., Pedrycz, W., Gacek, A., and Dai, Y. (2018). Development of information granules of higher type and their applications to granular models of time series. *Eng Appl of AI*, 71, 60–72.
- Lughofer, E. (2011). Evolving fuzzy systems: Methodologies, advanced concepts and applications. *Springer: Verlag Berlin Heidelberg*.
- Lughofer, E., Pratama, M., and Skrjanc, I. (2018). Incremental rule splitting in generalized evolving fuzzy systems for autonomous drift compensation. *IEEE Trans Fuzzy Syst*, 26(4), 1854–1865.
- Maciel, L., Ballini, R., and Gomide, F. (2016). Evolving granular analytics for interval time series forecasting. *Granular Computing*, 1(4), 213–224.
- Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism: principles and applications of bioelectric and bio-magnetic fields*. Oxford University Press, New York.
- Nguyen, H.L., Woon, Y.K., and Ng, W.K. (2015). A survey on data stream clustering and classification. *Knowl Inf Syst*, 45(3), 535–569.
- Ossmy, O., Tam, O., Puzis, R., Rokach, L., Inbar, O., and Elovici, Y. (2011). Minddesktop-computer accessibility for severely handicapped. *ICEIS*, 4, 316–320.
- Pedrycz, W., Al-Hmouz, R., Morfeq, A., and Balamash, A. (2013). The design of free structure granular mappings: the use of the principle of justifiable granularity. *IEEE Trans Cybern*, 43(6), 2105–2113.
- Pedrycz, W. and Homenda, W. (2013). Building the fundamentals of granular computing: A principle of justifiable granularity. *Appl Soft Comput*, 13(10), 4209–4218.
- Rösler, O. and Suendermann, D. (2013). First step towards eye state prediction using eeg. *Proc of the AIHLS*.
- Soares, E., Costa, P., Costa, B., and Leite, D. (2018). Ensemble of evolving data clouds and fuzzy models for weather time series prediction. *Appl Soft Comput*, 64, 445–453.
- Stigler, S.M. (1982). A modest proposal: a new standard for the normal. *Amer Statistician*, 36(2), 137–138.
- Tomažič, S., Dovžan, D., and Škrjanc, I. (2019). Confidence-interval fuzzy model-based indoor localization. *IEEE Trans Ind Electron*, 66(3), 2015–2024.
- Škrjanc, I., Iglesias, J., Sanchis, A., Leite, D., Lughofer, E., and Gomide, F. (2019). Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey. *Information Sciences*, 490, 344–368.
- Wang, X., Pedrycz, W., Gacek, A., and Liu, X. (2016). From numeric data to information granules: A design through clustering and the principle of justifiable granularity. *Knowl-Based Syst*, 101, 100–113.
- Yager, R. (1992). Default knowledge and measures of specificity. *Info Sci*, 61(1-2), 1–44.
- Yager, R. (2008). Measures of specificity over continuous spaces under similarity relations. *Fuzzy Sets Syst*, 159(17), 2193–2210.
- Yao, J., Vasilakos, A., and Pedrycz, W. (2013). Granular computing: Perspectives and challenges. *IEEE Trans Cybern*, 43(6), 1977–1989.
- Yao, Y. (2013). A triarchic theory of granular computing. *Granular Computing*, 1(2), 145–157.
- Zadeh, L.A. (1997). Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst*, 90(2), 111–127.
- Zadeh, L.A. (2006). Generalized theory of uncertainty (gtu) - principal concepts and ideas. *Comput Stat Data An*, 51(1), 15–46.
- Zdešar, A., Dovžan, D., and Škrjanc, I. (2014). Self-tuning of 2dof control based on evolving fuzzy model. *Appl Soft Comput*, 19, 403–418.