

Rule Exclusion Mechanism in Evolutionary Fuzzy Systems Mecanismo de Exclusão de Regras em Sistemas Fuzzy Evolutivos

Article Info:

Article history: Received 2022-10-25 / Accepted 2022-11-01 / Available online 2022-11-03

doi: 10.18540/jcecv18iss8pp14884-01e

Marcus Vinícius de Freitas Diadelmo

ORCID: <https://orcid.org/0000-0003-4855-1533>

Instituto Federal de Minas Gerais – Campus Avançado Itabirito, Brasil

E-mail: marcus.diadelmo@ifmg.edu.br

Arthur Caio Vargas e Pinto

ORCID: <https://orcid.org/0000-0002-0762-0072>

Instituto Federal de Minas Gerais – Campus Avançado Itabirito, Brasil

E-mail: arthur.vargas@ifmg.edu.br

Tamires Martins Rezende

ORCID: <https://orcid.org/0000-0002-5152-2488>

Machine Intelligence and Data Science Laboratory, Brasil

E-mail: inc.tamires@gmail.com

Resumo

O presente trabalho tem como objetivo propor um sistema de exclusão de regras e, conseqüentemente, a simplificação do modelo em sistemas fuzzy evolutivos. Tal simplificação tem alguns benefícios, podendo ser destacado, por exemplo, o trabalho de rotulação das regras por um especialista em sistemas não supervisionados e a explicação das regras obtidas. Para execução do trabalho foi considerado um algoritmo presente na literatura, ALMNo, com a adição do mecanismo de exclusão proposto. O mecanismo proposto utiliza a distância entre os centros das funções de pertinência das regras, normalizado pelo desvio padrão de uma janela deslizante com os últimos 10 dados analisados. A normalização visa detectar uma mudança no contexto dos dados, e, uma vez detectada a mudança, proporcionar uma maior generalização ao sistema. Isso se deve ao fato de que dados pertencentes a outra região do espaço gera um desvio padrão maior. Os resultados foram analisados comparando o algoritmo ALMNo original com o algoritmo ALMNo adicionado o mecanismo de exclusão. Resultados numéricos mostram que o mecanismo proposto é promissor, uma vez que reduziu o número de regras e manteve um nível competitivo de acurácia. Além disso, resultados de testes indicam que a definição dos parâmetros necessários não é algo decisivo para o sucesso do algoritmo.

Palavras-chave: Sistemas Fuzzy Evolutivos. Regras Fuzzy. Exclusão de regras.

Abstract

This paper aims to propose a rule exclusion system and, consequently, the model simplification in evolutionary fuzzy systems. Such simplification has some benefits, being highlighted, for example, the task of labelling the rules by an expert in unsupervised systems and the explanation of the rules obtained. For the execution of the work, it was considered an algorithm present in the literature, ALMNo, with the addition of the proposed exclusion mechanism. The proposed mechanism uses the distance between the centers of the membership functions of the rules, normalized by the standard deviation of a sliding window with the last 10 data analyzed. The normalization is intended to detect a change in the context of the data, and once it is detected, provides greater generalizability

to the system. This is due to the fact that data belonging to another region of space generates a larger standard deviation. The results were analyzed by comparing the original ALMNo algorithm with that without the exclusion mechanism. Numerical results show that the proposed mechanism is promising in terms of reducing the number of rules and maintaining a competitive level of accuracy. Furthermore, test results indicate that setting the necessary parameters is not decisive for the success of the algorithm.

Keywords: Evolutionary Fuzzy Systems. Fuzzy Rules. Rule Exclusion.

1. Introdução

Em problemas envolvendo data stream, a utilização de sistemas online, ou evolutivos, se torna uma poderosa ferramenta (Buswell *et al.*, 2001). Isso se deve ao fato de que sistemas online observam um conjunto de dados para gerar um modelo e fazer previsões. Contudo, a chegada de novos dados faz com que o sistema se adapte de forma autônoma ao novo contexto, sem perder informações importantes já aprendidas (Škrjanc *et al.*, 2019). Narendra *et al.* (1995) destacam que estruturas baseadas em múltiplos modelos, como os sistemas baseados em regras fuzzy, são muito eficientes para tratar problemas online.

Em Škrjanc *et al.* (2019) é apresentada uma gama de trabalhos na linha de sistemas evolutivos. Em um dos trabalhos apresentados, base para o atual trabalho, é proposta a obtenção de estruturas baseadas em agrupamentos (Angelov *et al.*, 2017). Os autores destacam que o algoritmo proposto não requer nenhuma parametrização *a priori*.

Apesar dos algoritmos não necessitarem mais de um especialista para auxiliar o treinamento, a utilização de um pode complementar tais sistemas, identificando e interpretando os resultados de um sistema não supervisionado (Wang e Mendel, 1992). Além disso, atualmente a linha de Inteligência Artificial entra no ramo de explicar suas decisões e não somente entregar resultados em caixas pretas (Murray *et al.*, 2021). Dessa forma, o auxílio de um especialista contribuiria para tal tarefa.

O objetivo do trabalho visa reduzir a complexidade de um sistema evolutivo no que diz respeito ao número de regras, tal redução viabiliza, por exemplo, a identificação do conseqüente das regras por um especialista. Com um número menor de regras o trabalho de rotulação se torna menos complexo. Contudo, tal redução só é viável caso o desempenho de classificação do algoritmo se mantenha eficiente. Dessa forma, o trabalho propõe uma técnica de redução de regras com melhoria ou manutenção de desempenho. Para tal tarefa, foi utilizado o sistema evolutivo ALMMo, proposto por Angelov *et al.* (2017), mas com a adição de um mecanismo de exclusão de regras.

O mecanismo de exclusão de regras visa detectar mudanças de estruturas, o qual pode indicar a presença de agrupamentos e, então, definir regras específicas para os agrupamentos. A Figura 1 exemplifica essa detecção. Ao olhar somente para a Figura 1.a, alguns analistas podem afirmar que somente um agrupamento pode ser identificado, enquanto outros podem indicar que existam mais. Contudo, ao analisar a Figura 1.b, possivelmente é consenso que o sistema possua dois agrupamentos. Dessa forma, a motivação do trabalho é detectar essa variação da estrutura e tornar o sistema mais generalista.

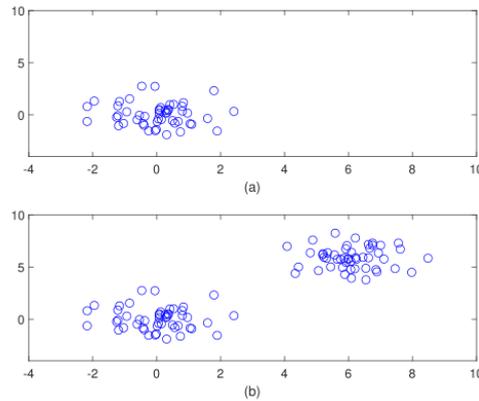


Figura 1. Mudança de contexto. (a) Conjunto de dados pode indicar a presença de vários agrupamentos. (b) Conjunto de dados indica dois agrupamentos.

2. Sistema Evolutivo ALMMo

O algoritmo proposto por Angelov *et al.* (2017) apresenta o mecanismo evolutivo baseado em dois estágios (identificação da estrutura e identificação de parâmetros). O sistema é baseado em regras *fuzzy*, onde os antecedentes das regras são agrupamentos e os consequentes são pesos de equações lineares. A Figura 2 e as Equações 1 e 2 mostram este esquema.

$$SE \ (x \sim \Xi_j) \ ENTÃO \ (y_j = \mathbf{u}^T \mathbf{a}_j) \quad (1)$$

$$y = \sum_{j=1}^N \lambda_j y_j \quad (2)$$

Sendo Ξ a j^{esima} regra, y_j a saída da j^{esima} regra, $\mathbf{u}^T = [1, \mathbf{x}^T]$, \mathbf{a}_j os parâmetros consequentes da j^{esima} regra, λ_j o nível de ativação da j^{esima} regra e, finalmente, y é a saída do modelo.

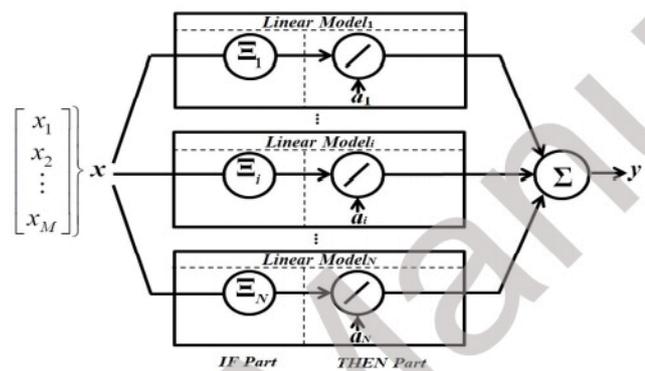


Figura 2. Sistema proposto por Angelov *et al.* (2017). Adaptado de Angelov *et al.* (2017).

Os autores iniciam o sistema identificando a existência de novas estruturas, baseadas na densidade do conjunto de dados. Com a chegada de um novo dado, é verificado se este possui uma

densidade fora da Densidade global, caso isso ocorra, será gerada uma nova possível regra em torno do novo dado. A densidade da amostra x é calculada de acordo com a Equação 3.

$$D_K(x) = \frac{1}{1 + \frac{||x - \mu_K||^2}{\sigma_K^2}} \quad (3)$$

onde $\sigma_K^2 = X_K - ||\mu_K||^2$, μ_K e X_K são a média e produto escalar médio do conjunto de dados. Como o algoritmo é evolutivo, tais valores são atualizados pelas Equações 4 e 5.

$$\mu_K = \frac{K-1}{K} \mu_{K-1} + \frac{1}{K} x_K; \mu_1 = x_1 \quad (4)$$

$$X_K = \frac{K-1}{K} X_{K-1} + \frac{1}{K} ||x_K||^2; X_1 = ||x_1||^2 \quad (5)$$

sendo X_K a amostra na iteração K .

O sistema ainda verifica se a nova possível regra sobrepõe alguma já existente, caso isso ocorra, a regra já existente é atualizada por valores médios entre a regra existente e a nova possível regra. Ou seja, dessa forma não foi gerada uma nova regra, mas sim atualizada uma já existente.

Caso a nova possível regra não sobreponha nenhuma outra, a amostra x será a nova regra.

No caso da nova amostra x não representar uma nova possível regra, está será direcionada à regra mais próxima, encontrando o argumento mínimo $||x - \mu_i||$ em que i varia de 1 até o total de regras. E então, os parâmetros do grupo são atualizados pelas Equações 6 e 7.

$$\mu_{K+1} = \frac{S}{S+1} \mu_K + \frac{1}{S+1} x \quad (6)$$

$$X_{K+1} = \frac{S}{S+1} X_{K-1} + \frac{1}{S+1} ||x||^2 \quad (7)$$

onde S representa a quantidade de amostras associadas ao grupo. Vale destacar que as atualizações dadas pelas Equações 6 e 7 só acontecem no grupo ao qual a nova amostra está associada.

Após a geração dos parâmetros acima, torna-se necessário atualizar os parâmetros consequentes das regras. Caso uma nova regra seja adicionada é necessário inicializar os parâmetros consequentes de acordo com as Equações 8 e 9.

$$\mathbf{a}_{K,N_{K+1}} = \frac{1}{N_K} \sum_{j=1}^{N_K} \mathbf{a}_{K,j} \quad (8)$$

$$\mathbf{C}_{K,N_{K+1}} = \Omega \mathbf{I} \quad (9)$$

sendo N_K o número de regras, Ω um valor constante, \mathbf{I} uma matriz identidade de dimensão $\mathbf{G} + 1$, sendo \mathbf{G} a dimensão do problema de entrada. No caso de uma regra substituir uma antiga, a nova regra herdar os parâmetros da antiga.

Finalmente os parâmetros dos consequentes podem ser atualizados pelas Equações 10 e 11.

$$\mathbf{C}_{K+1,j} = \mathbf{C}_{K,j} - \frac{\lambda_{K+1,j} \mathbf{C}_{K,j} \mathbf{u}_{K+1}^T \mathbf{u}_{K+1}^T \mathbf{C}_{K,j}}{1 + \lambda_{K+1,j} \mathbf{u}_{K+1}^T \mathbf{C}_{K,j} \mathbf{u}_{K+1}} \quad (10)$$

$$\mathbf{a}_{K+1,j} = \mathbf{a}_{K,j} + \lambda_{K+1,j} \mathbf{C}_{K+1,j} \mathbf{u}_{K+1} (y_{K+1} \mathbf{u}_{K+1}^T \mathbf{a}_{K,j}) \quad (11)$$

Angelov *et al.* (2017) apresentam outras técnicas para regressão e aprendizado *offline*. Entretanto, como a proposta implementada visa o aprendizado online, as outras técnicas não serão apresentadas.

2. Metodologia Proposta

Este trabalho propõe uma extensão ao trabalho proposto por Angelov *et al.* (2017). Onde a proposta visa a exclusão de regras baseado na análise da distância dos centros ponderados pela variância de uma janela deslizante dos últimos dados. Tal ideia propõe que a presença de novas regras deva gerar uma maior generalização do modelo.

A normalização dos dados visa detectar mudanças de contexto dos dados e permitir uma melhor distribuição das regras. A condição para excluir regras é dada pela Equação 12.

$$SE(dcmin < \alpha) ENTÃO(excluaaregra) \quad (12)$$

em que α é a tolerância para excluir novas regras, um valor maior indica mais regras excluídas; $dcmin$ (Equação 14), é a distância entre dois centros dos antecedentes das regras dividido por um valor de normalização, denominado $norm$ (Equação 13); que é a variância entre os últimos 10 dados analisados. É importante destacar que a análise da condição descrita prioriza deletar as regras mais antigas. Ou seja, primeiro a regra mais nova é comparada à regra mais antiga. Dessa forma, a regra mais nova nunca será deletada.

Após a chegada de um novo dado (x_K) e a criação de uma nova regra, o sistema de exclusão é acionado para avaliar se alguma regra deve ser excluída. O primeiro passo é calcular o valor de $norm$, dado pela equação 13.

$$norm = \max[DV(J_x)] \quad (13)$$

sendo $DV(\cdot)$ o desvio padrão e J_x um vetor com os últimos 10 dados analisados. Para K menor que 10, J_x é o vetor dos dados de 1 até K . É importante destacar que no momento em que o contexto dos dados se altera o valor de $norm$ aumenta, fazendo com que o valor de $dcmin$ reduza e permita um maior número de regras excluídas.

Após o cálculo de $norm$, a distância normalizada entre os centros dos antecedentes das regras é calculada pela Equação 14.

$$dcmin = \frac{\|C_j - C_i\|}{norm} \quad (14)$$

sendo $C_{()}$ as coordenadas do antecedente da regra, $j = N, N - 1, \dots, 1$ e $i = 1, 2, \dots, N - 1$. No momento em que $dcmin < tol$ a regra i é imediatamente eliminada. Dessa forma, regras mais antigas possuem maiores chances de serem excluídas. O Algoritmo 1 detalha o mecanismo já descrito.

Algoritmo 1. Remove Regras

```

N ← Número de regras
norm ← max[DV(Jx)]
ex ← []
for j ← N to 2 do
  for i ← 1 to N - 1 do
    dmin ← ||Cj - Ci|| / norm
    z ← posição onde (ex = 1)
    if i = j ou z ≠ vazio then
      dmin ← tol + 1
    end if
    if dmin < α then
      Remova regra i
      ex ← [ex; i]
    end if
  end for
end for

```

As Figuras 3 e 4 mostram os centros dos consequentes para os casos em que foi e não foi utilizado o mecanismo para deletar regras. Nota-se que a Figura 3 apresenta uma generalização maior.

Pela Figura 5 é possível notar que, enquanto os dados de outro agrupamento não chegam, o sistema permite mais regras próximas. A Figura 5.a mostra que como os dados da outra classe ainda são poucos, o sistema permitiu duas regras para uma das classes. Já a Figura 5.b mostra que a chegada de novos dados gerou um novo agrupamento e, como consequência, o sistema excluiu uma das regras existentes.

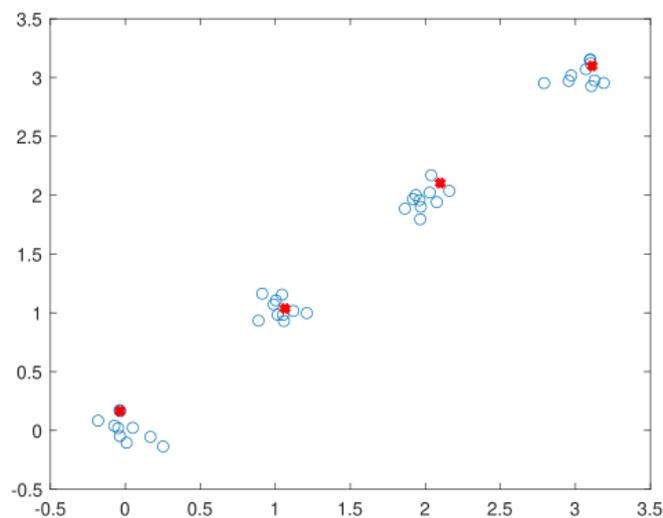


Figura 3. Modelo Final após treinamento online com mecanismo para deletar regras. Centros das regras em vermelho.

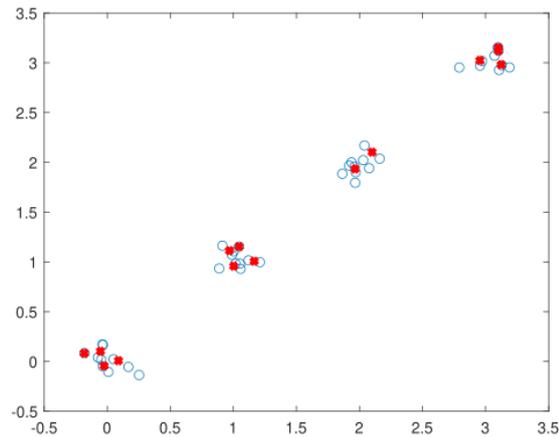


Figura 4. Modelo Final após treinamento online sem mecanismo para deletar regras. Centros das regras em vermelho.

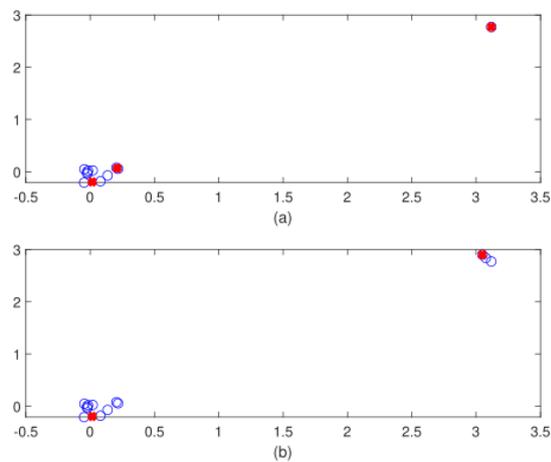


Figura 5. Chegada e geração de novos clusters e, exclusão de regras. (a) novo cluster sendo detectado, (b) novo cluster detectado.

3. Resultados e Discussão

Para avaliar a viabilidade do algoritmo foram utilizadas, além das duas bases utilizadas em Angelov et al. (2017), outras três bases presentes na literatura que podem ser vistas na Tabela 1. As bases possuem tamanhos, atributos e características distintas, contudo, todas as bases são de classificação binária.

1.1 Desempenho do modelo proposto

Para avaliar a viabilidade do algoritmo foram utilizadas, além das duas bases utilizadas em Angelov et al. (2017), outras três bases presentes na literatura que podem ser vistas na Tabela 1. As bases possuem tamanhos, atributos e características distintas, contudo, todas as bases são de classificação binária.

Tabela 1. Bases de dados atualizadas.

Base	Treino/Teste	Atributos
Bank	1235/137	4
Diabetes	691/77	8
ILPD	521/58	10
Occupancy	8143/975	5
Sonar	187/21	60

Como forma de avaliar o desempenho da técnica proposta, os dados foram divididos em um conjunto de treinamento, composto por 90% do total de dados, e um conjunto de teste, com os 10% restantes, como mostrado na Tabela 1. Foram realizadas 30 execuções para cada algoritmo, e para cada execução, os dados foram embaralhados antes de se realizar a separação entre treinamento e teste. Para todas as bases o parâmetro α foi sempre igual a 10.

A Tabela 2 apresenta a acurácia média, entre as 30 execuções, encontrada para o método proposto e o algoritmo ALMMo. O cálculo de acurácia é mostrado na Equação 15. Além disso, é mostrado o número de regras (classes) para cada base. Apesar do número de regras ser sempre um número inteiro, a tabela apresenta valores fracionados pois apresenta o número médio de regras das 30 execuções.

$$acuracia = \frac{\#acertos}{\#dadosteste} \quad (15)$$

Tabela 2. Desempenhos dos métodos. Acurácia média e número de regras média após 30 execuções.

Base	Proposta	ALMNo
Bank	0,9759/2	0,9779/22,9
Diabetes	0,7576/9,3	0,7532/16,4
ILPD	0,7006/3,7	0,6989/15,7
Occupancy	0,9902/16,3	0,9901/28,7
Sonar	0,7476/7,3	0,7508/7,5

onde $\#acertos$ representa o número de dados de teste rotulados corretamente e $\#dadosteste$ o número total de dados de teste.

Os resultados mostram que a metodologia simplifica o modelo, pois apresentou menos regras em todos os casos, e que não impacta, necessariamente, em uma piora de desempenho de classificação.

1.2 Influência do parâmetro α

Como a metodologia proposta possui um parâmetro que deve ser ajustado *a priori* pelo usuário, foi realizado um experimento para verificar qual o impacto da mudança deste parâmetro no desempenho geral do classificador. Para tal, foi realizada uma rodada do algoritmo para cada base e para cada valor de α . Para α foram considerados os valores (0.001, 0.01, 0.1, 1, 10, 20, 30, 40, 50, 100).

A Figura 6 mostra o experimento realizado. Apesar de não ter sido realizado um teste estatístico é possível notar que o aumento de α implica em mais regras sendo excluídas, resultado

esperado, uma vez que um maior valor representa maiores chances de exclusão de alguma regra. Com relação à acurácia, não é possível inferir um padrão. Para alguns casos a acurácia melhorou com o aumento de α , para outras piorou e para outras não teve grande impacto, indicando que, possivelmente, a mudança de α não influencia, significativamente, no desempenho do classificador.

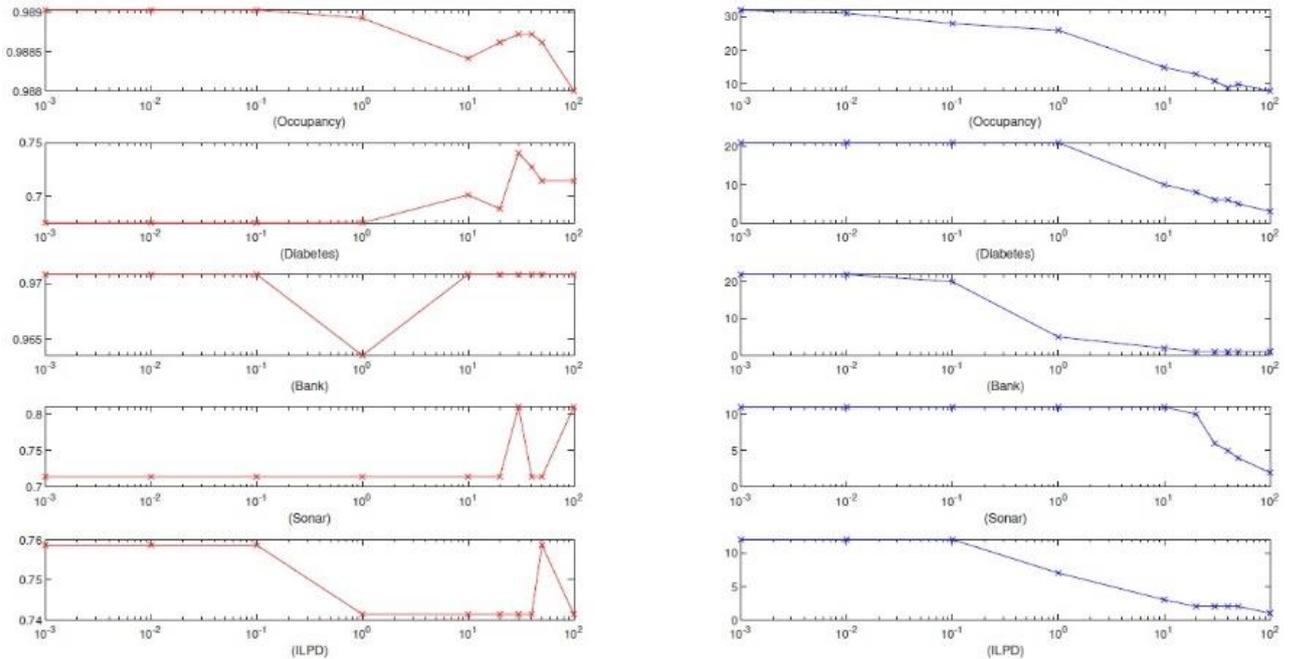


Figura 6. Acurácia e número de regras ao variar α . O eixo x dos gráficos indica o valor de α , escala logarítmica. Os gráficos da esquerda, em vermelho, indicam a acurácia e os da direita, em azul, indicam o número de regras.

4. Conclusão

Com os resultados obtidos, notou-se que o sistema proposto reduziu o número de regras e, como consequência, reduziu a complexidade do sistema. Como dito, a redução do número de regras influencia em uma possível rotulação por um especialista, reduzindo seu trabalho e, podendo facilitar a explicação dos resultados obtidos.

Porém, tal resultado só torna o sistema viável pois o desempenho do sistema em classificar dados se mostrou eficiente, em comparação ao método original. Os resultados mostram que houve redução do número de regras e a acurácia se mostrou competitiva com o método original.

O método proposto depende de um parâmetro de tolerância definido a priori. Os testes realizados mostraram que tal parâmetro não altera significativamente o desempenho do sistema, deixando o sistema menos dependente de conhecimentos específicos para definição do parâmetro de tolerância. Tal análise sugere, ainda, que o parâmetro possa ser ajustado dinamicamente durante o processo online, caso o usuário queira tentar reduzir ou aumentar o número de regras do sistema.

Referências

- Angelov, P. P., Gu, X., & Principe, J. C. (2017). Autonomous learning multimodel systems from data streams. *IEEE Transactions on Fuzzy Systems*, 26(4), 2213-2224.
- Buswell, R., Angelov, P., & Wright, J. (2001). Transparency and simplification of rule-based models for on-line adaptation. *Eusflat*.
- Gámez, J. A., & Puerta, J. M. (2009). Learning weighted linguistic fuzzy rules by using specifically-tailored hybrid estimation of distribution algorithms. *International journal of approximate reasoning*, 50(3), 541-560.
- Murray, B. J., Islam, M. A., Pinar, A. J., Anderson, D. T., Scott, G. J., Havens, T. C., & Keller, J. M. (2020). Explainable ai for the choquet integral. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(4), 520-529.
- Narendra, K. S., Balakrishnan, J., & Ciliz, M. K. (1995). Adaptation and learning using multiple models, switching, and tuning. *IEEE control systems magazine*, 15(3), 37-51.
- Škrjanc, I., Iglesias, J. A., Sanchis, A., Leite, D., Lughofer, E., & Gomide, F. (2019). Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey. *Information Sciences*, 490, 344-368.
- Wang, L. X., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on systems, man, and cybernetics*, 22(6), 1414-1427.