



**UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA**

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76  
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA**

## **XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS** **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2022**

### **CONSTRUÇÃO GENÉTICA MULTIOBJETIVO DE CLASSIFICADORES** **FUZZY PARA FLUXO DE DADOS**

**Allan Pereira da Silva<sup>1</sup> e Matheus Giovanni Pires<sup>2</sup>**

1. Bolsista PIBIC/FAPESB, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: [allanpereira016@gmail.com](mailto:allanpereira016@gmail.com)
2. Orientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, e-mail: [mgpires@comp.uefs.br](mailto:mgpires@comp.uefs.br)

**PALAVRAS-CHAVE:** sistemas de classificação fuzzy; aprendizado genético de sistemas fuzzy; fluxo de dados.

#### **INTRODUÇÃO**

A extração de conhecimento através de uma base de dados é extremamente relevante em diversos campos de pesquisa, tais como, mineração de dados e aprendizado de máquina. Além disso, grandes setores comerciais e industriais passaram a utilizar o poder dos dados para auxiliar na tomada de decisão. Normalmente, a aprendizagem é realizada em ambientes estacionários, possibilitando uma alta disponibilidade dos dados para qualquer instante de tempo sem que os mesmos sofram variações em relação ao seu objetivo de classificação. Na literatura, diversos autores encontraram soluções satisfatórias para classificação de dados estacionários (Martins et al., 2009; Cattani et al., 2013).

Entretanto, o aprendizado de uma base não estacionária não possui uma literatura tão abrangente em comparação as bases estacionárias. Além de ser uma área do conhecimento mais recente, a complexidade de extrair conhecimento em ambientes de fluxo contínuo é elevada. Alguns exemplos estão interligados na análise de fluxo contínuo de dados, dentre eles: detecção de presença por meio de sensores; monitoramento da temperatura da água; sistemas de recomendação; preferência de roupas em relação a estação; preferências por notícias; consumo de energia e filtragem de spam. Considerando o exemplo da filtragem de spam, ele é considerado um problema não estacionário devido à volatilidade dos filtros que classificam de forma incorreta os e-mails.

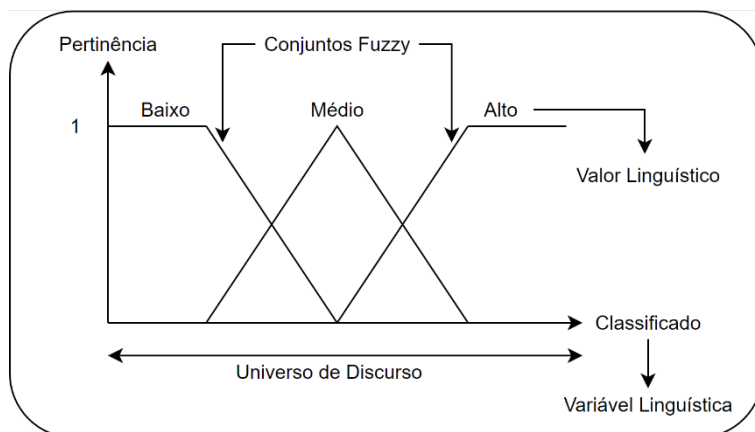
Comparando com a abordagem tradicional de aprendizado, o processamento dos dados em ambientes com fluxo contínuo normalmente impõe restrições em relação ao uso da memória, tempo de execução e quantidade de instâncias disponíveis no treinamento (SANTOS, 2015). Além disso, algumas dificuldades são encontradas neste contexto. Geralmente os classificadores perdem eficácia ao longo do tempo, o que ocasiona uma instabilidade em relação a sua taxa assertiva (Gama, 2010) e a mudança na distribuição dos dados é responsável pelos problemas conhecidos como desvios de conceito (*concept drift*).

## MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)

Para a definição dos conjuntos fuzzy e da base de regras do sistema fuzzy foi necessário a utilização de uma base de dados para cada problema a ser tratado, onde cada um dos exemplos é composto por atributos de entrada e saída.

Em relação à codificação dos conjuntos fuzzy foram utilizados três conjuntos para representar cada variável de entrada, sendo dois em formato trapezoidal e um central, em formato triangular, como pode ser visto na Figura 1.

**Figura 1:** Conjuntos fuzzy.



A base de regras foi gerada pelo algoritmo de Wang-Mendel (Fan et al., 2019), a partir dos atributos pertencentes aos exemplos de cada base de dados. O algoritmo de Wang-Mendel possui facilidade de implementação, sendo capaz de gerar regras com bom poder de classificação.

O método de raciocínio clássico foi o mecanismo de inferência utilizado para classificar os padrões de acordo com a base de regras do sistema fuzzy. Este método agrega os graus de compatibilidade de todas as regras em relação ao exemplo de entrada e o classifica utilizando uma disputa entre as regras. A regra que possui o maior grau de compatibilidade em relação ao exemplo de entrada é escolhida para classificá-lo.

A codificação dos cromossomos é do tipo inteiro, onde cada antecedente das regras é representado por um número entre um e três e o consequente da regra irá variar de acordo com o número de classes de cada problema. Cada indivíduo da população representa, então, uma base de regras do classificador fuzzy, ou seja, uma solução para o problema.

A população inicial é gerada a partir de uma delimitação fornecida por uma semente. Esta semente fornece um conhecimento prévio sobre o problema em questão e sua composição armazena a base de regras gerada pelo algoritmo de Wang-Mendel. A necessidade de utilizar uma semente surgiu para reduzir o tempo de convergência e minimizar os recursos computacionais necessários, como alocação de memória e processamento.

O tamanho da população foi definido em 100 indivíduos, onde o primeiro é a semente e os demais gerados aleatoriamente, os quais foram delimitados pela semente. Os indivíduos gerados aleatoriamente estão ligados diretamente a composição de regras.

Cada indivíduo da população é avaliado de acordo com sua taxa de acurácia e interpretabilidade. A acurácia leva em consideração a precisão do classificador na

classificação dos exemplos e a interpretabilidade considera a redução da base de regras, a qual é otimizada por um Algoritmo Genético Multiobjetivo.

Na literatura existem vários algoritmos para detecção de desvios. Barros e Santos (2018) comparam 14 detectores de desvios de conceito em larga escala, variando as características fundamentais de cada base de dados. De acordo com Barros e Santos (2018), o algoritmo *Drift Detection Methods based on Hoeffding's Bounds* (HDDM) nas versões A-Test e W-Test se destacam pela eficiência e precisão na detecção de desvios de conceito. HDDM A-Test e HDDM W-Test são utilizados para detectar desvios abruptos e graduais, respectivamente.

Para a construção genética do classificador fuzzy, os métodos HDDM A-Test e HDDM W-Test serão acoplados após o resultado da classificação do fluxo de dados, com o objetivo de sinalizar a imprecisão do classificador ao longo do tempo, ou seja, serão responsáveis pela sinalização da necessidade de retreinar o classificador durante o recebimento do fluxo de dados.

Inicialmente, a entrada é um fluxo teoricamente infinito de dados que serão classificados de acordo com um classificador base. Enquanto o detector de desvio não sinalizar uma variação nos dados de entrada, o mesmo classificador base será utilizado. O classificador base só será alterado quando um desvio for sinalizado pelos detectores de desvio.

### **RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)**

Primeiramente, vale destacar que as atividades deste plano de trabalho foram realizadas no período de 01/10/2021 a 30/03/2022, ou seja, ainda restavam seis meses para cumprir com todas as atividades planejadas. Mas, como o prazo para substituição do bolsista foi expirado, não foi possível dar continuidade neste trabalho.

Todas as atividades planejadas foram executadas até este momento, sendo assim, os principais resultados obtidos estão a implementação de um Sistema de Classificação Baseado em Regras Fuzzy, a implementação de um Algoritmo Evolutivo Multiobjetivo e o processo de leitura dos conjuntos de dados.

Sendo assim, de acordo com as próximas atividades planejadas, os testes iniciariam para obter os primeiros resultados da solução proposta. A partir destes resultados, ajustes no algoritmo evolutivo deveriam ser realizados, com o objetivo de melhorar cada vez mais os resultados de precisão do sistema de classificação.

### **CONSIDERAÇÕES FINAIS (ou Conclusão)**

As atividades desenvolvidas neste projeto permitiram ao aluno adquirir habilidades de programação, conhecimentos mais profundos em Inteligência Artificial, com ênfase em classificação de fluxos de dados contínuos, sistemas fuzzy e algoritmos genéticos multiobjetivos. Esta grande experiência contribui fortemente para a formação do aluno como futuro Engenheiro de Computação.

### **REFERÊNCIAS**

CATTANI, C. E. V., MERCANTE, E., SOUZA, C., E WRUBLACK, S. C. (2013). Desempenho de algoritmos de classificação supervisionada para imagens dos satélites *rapideye*. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto-SBSR, páginas 8005–8010.

GAMA, J. (2010). Knowledge discovery from data streams. CRC Press.

MARTINS, A. C., MARQUES, J. M., E COSTA, P. D. (2009). Estudo comparativo de três algoritmos de *machine learning* na classificação de dados electrocardiográficos. Trabalho (Mestrado em Informática Médica), Universidade do Porto, Porto.

SANTOS, S. G. T. d. C. (2015). Avaliação criteriosa dos algoritmos de detecção de *concept drifts*. Dissertação de Mestrado, Universidade Federal de Pernambuco.

FAN, Z., CHIONG, R., HU, Z., DHAKAL, S., E LIN, Y. (2019). A two-layer wang-mendel fuzzy approach for predicting the residuary resistance of sailing yachts. *Journal of Intelligent & Fuzzy Systems*, 36(6):6219–6229.

BARROS, R. S. M. E SANTOS, S. G. T. C. (2018). A large-scale comparison of concept drift detectors. *Information Sciences*, 451:348–370.