

Leonardo Santos Paulucio

## **Tarefa Computacional 2**

Vitória - ES

15 de Novembro de 2019

# 1 Introdução

O objetivo desse trabalho computacional é realizar a classificação de um conjunto de dados usando lógica *fuzzy* combinadas com algoritmos evolutivos. O trabalho é baseado no trabalho desenvolvido por [Svensson \(2019\)](#).

O trabalho é composto de duas partes. A primeira corresponde na definição das variáveis linguísticas e seus respectivos conjuntos *fuzzy*, da base de regras e do classificador *fuzzy*. Já a segunda parte consiste em achar as funções de pertinências e a base de regra utilizando os algoritmos evolutivos PSO e GA. A *fitness* utilizada nesse sistema será o erro de acurácia do classificador em todo o *dataset* da flor Íris. A classe para qual o dado de entrada será classificado irá ser aquela que possuir o maior grau de pertinência entre as três classes.

O *dataset* Íris consiste de 150 amostras rotuladas, sendo 50 para cada uma das três espécies: Setosa, Versicolor e Virginica. Cada amostra possui informações sobre 4 parâmetros: tamanho da sépala, comprimento da sépala, tamanho da pétala e comprimento da pétala.

## 2 Sistema Fuzzy

### 2.1 Variáveis Linguísticas

As variáveis linguísticas utilizadas foram: *short*, *middle* e *long*, ilustrado na Figura 1. Nesse trabalho o valor do parâmetro  $W$  será calculado utilizando os algoritmos evolutivos PSO e GA. Dessa forma, para cada um dos 4 parâmetros de entrada um valor  $W_i$  correspondente será definido, onde  $i$  corresponde ao número da *feature*. Com esses valores será medido a acurácia do modelo em todo o *dataset*.

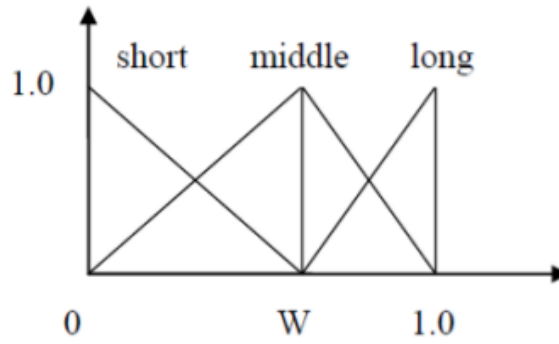


Figura 1 – Grau de pertinência das variáveis *short*, *middle* e *long*

### 2.2 Base de Regras

Para a base de regras foram utilizadas as 4 regras descritas abaixo.

$R_1$ : SE  $X_1 = (\text{short OU long})$  E  $X_2 = (\text{middle OU long})$  E  $X_3 = (\text{middle OU long})$  E  $X_4 = \text{middle}$ :

ENTÃO iris = versicolor

$R_2$ : SE  $X_3 = (\text{short OU middle})$  E  $X_4 = \text{short}$ :

ENTÃO iris = setosa

$R_3$ : SE  $X_2 = (\text{short OU middle})$  E  $X_3 = \text{long}$ :

ENTÃO iris = versicolor

$R_4$ : SE  $X_1 = \text{middle}$  E  $X_2 = (\text{short OU middle})$  E  $X_3 = \text{short}$  E  $X_4 = \text{long}$ :

ENTÃO iris = versicolor

Onde  $X_1 = \text{sepal legth}$ ,  $X_2 = \text{sepal width}$ ,  $X_3 = \text{petal legth}$  e  $X_4 = \text{petal width}$ .

## 3 Resultados e Discussão

Para a realização dos testes foram definidos os parâmetros listados na Tabela 1. Os dados de entrada do *dataset* também foram normalizados para o intervalo de  $[0, 1]$ .

Parâmetro	Valor	Significado
NumGerações	50	Número de gerações
PopSize	20	Tamanho da População
NumRuns	20	Número de execuções
CrossProb	0.9	Prob. de Crossover
MutProb	0.1	Prob. de Mutação
$\chi$	0.7298	Constante de inércia
c1	2.05	Constante cognitiva
c2	2.05	Constante social

Tabela 1 – Parâmetros utilizados

### 3.1 *Baseline*

Foi feito um *baseline* utilizando um valor  $W_i = 0.5$  para todas as *features*. Os resultados obtidos estão ilustrados na Tabela 2. E a matriz de confusão obtida está ilustrada na Figura 2.

Dataset	Acurácia
Íris	0.86

Tabela 2 – Resultado da acurácia média obtida no baseline

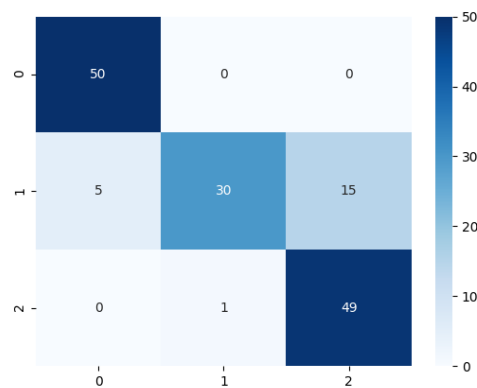


Figura 2 – Matriz de confusão obtida no baseline

## 3.2 Experimento

As tabelas seguintes apresentam os resultados obtidos no experimento. A Figura 3 apresenta as matrizes de confusão obtidas para os melhores  $W$ .

Métrica	Melhor	Média	Pior
Acurácia	0.97	0.95	0.85
Tempo(s)	36.58	37.24	38.28

Tabela 3 – Resultados obtidos com GA

Métrica	Melhor	Média	Pior
Acurácia	0.98	0.94	0.68
Tempo(s)	21.17	25.84	42.81

Tabela 4 – Resultados obtidos com PSO

Classificador	$W_1$	$W_2$	$W_3$	$W_4$
Fuzzy+GA	0.07	0.34	0.48	0.26
Fuzzy+PSO	0.02	0.26	0.43	0.27

Tabela 5 – Melhores  $W$  obtidos

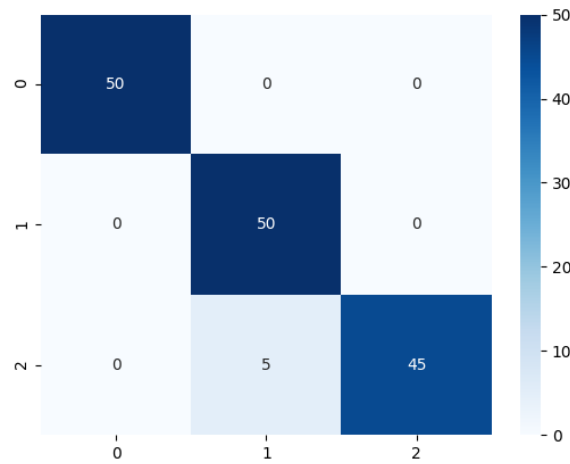
Classificador	Acurácia
Baseline	0.86
Fuzzy+GA	0.97
Fuzzy+PSO	<b>0.98</b>
Original	0.95

Tabela 6 – Comparação das melhores acurácias

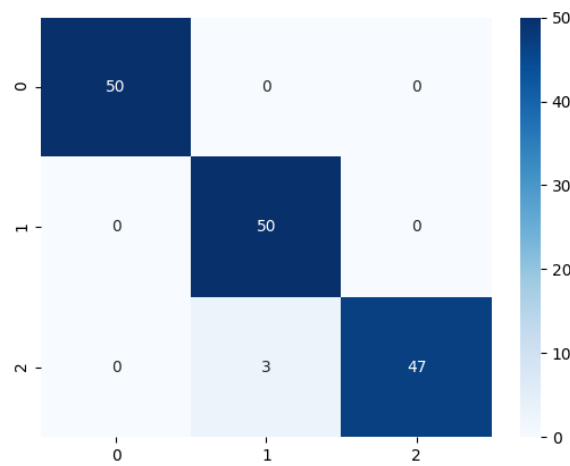
## 3.3 Discussões

Analisando-se as tabelas é possível notar que o GA obteve uma acurácia média maior, porém o PSO conseguiu obter a melhor acurácia nos testes chegando a 98%, superando até mesmo o obtido no artigo original. Com relação ao tempo de execução o PSO foi, em média, mais rápido, mas ainda assim ambos os algoritmos se mostraram bem rápidos.

Ao se comparar as matrizes de confusão da Figura 3 com a matriz de confusão obtida no *baseline* nota-se que os erros de classificação diminuem consideravelmente. Analisando-se as matrizes de confusão obtidas com Fuzzy+GA e Fuzzy+PSO percebe-se que são muito semelhantes, sendo que o erro ocorrido em ambas está na classificação de flores da classe *Vírginica* (1) como sendo da classe *Versicolor* (2), o que mostra uma certa dificuldade em se classificar essas duas classes.



(a) GA



(b) PSO

Figura 3 – Matrizes de confusão obtidas para as melhores acurácias

Assim, percebe-se que o uso de sistemas fuzzy com algoritmos evolutivos para sintonia das funções de pertinência apresentaram bons resultados, o que mostra serem boas opções para se conseguir otimizar classificadores fuzzy de forma automática.

# Referências

SVENSSON, S. *Implementing a Fuzzy Classifier and Improving its Accuracy using Genetic Algorithms*. [S.l.]: unpublished, 2019. Citado na página [1](#).