

UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG  
CENTRO DE CIÊNCIAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO  
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

**Utilizando funções de (pré)-agregação derivadas da  
Integral de Choquet em Sistemas de Classificação  
Baseados em Regras *Fuzzy***

Paula Fernanda Schiavo

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande - FURG, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação

Orientador: Prof. Dr. Graçaliz Pereira Dimuro  
Co-orientador: Prof. Dr. Eduardo Nunes Borges  
Colaborador: Prof. Dr. Humberto Bustince

Rio Grande, 2019

## Ficha catalográfica

S329u Schiavo, Paula Fernanda.

Utilizando funções de (pré)-agregação derivadas da integral de Choquet em sistemas de classificação baseados em regras *fuzzi* / Paula Fernanda Schiavo. – 2019.

55 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Computação, Rio Grande/RS, 2019.

Orientadora: Dra. Graçaliz Pereira Dimuro.

Coorientador: Dr. Eduardo Nunes Borges.

Colaborador: Dr. Humberto Bustince.

1. Sistemas baseados em regras *fuzzi* 2. Raciocínio aproximado 3. Integral Choquet 4. Funções de *overlap* 5. Índices de *overlap* 6. Funções de pré-agregação I. Dimuro, Graçaliz Pereira II. Borges, Eduardo Nunes III. Bustince, Humberto IV. Título.

CDU 004.021



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO RIO GRANDE  
CENTRO DE CIÊNCIAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO  
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

**Utilizando funções de (pré)-agregação derivadas da Integral de Choquet  
em Sistemas de Classificação Baseados em Regras Fuzzy**

Paula Fernanda Schiavo

Banca examinadora:

---

Prof. Dr. Benjamin René Callejas Bedregal

---

Prof. Dr. Eder Mateus Nunes Gonçalves

---

Prof. Dr. Héliida Salles Santos

---

Prof. Dr. Graçaliz Dimuro Pereira  
Orientador(a)

---

Prof. Dr. Eduardo Nunes Borges  
Coorientador(a)

## **AGRADECIMENTOS**

Agradeço a oportunidade de trabalhar com um excelente corpo docente que me proporcionou grande conhecimento na área que eu almejei. Em especial a professora doutora Graçaliz Dimuro que me apresentou um nova perspectiva de vida com o seu amplo conhecimento e ao professor doutor Eduardo Borges pela dedicação no ensino.

Aos meus pais por terem me passado valores e uma mensagem essencial para a vida: “o estudo como herança”.

Ao meu companheiro Sidnei pela caminhada e sua forma de ver a vida, admiro-o.

Aos meus sogros, Sidnei e Lúcia, por serem parte ativos no dia-a-dia e na conclusão desta dissertação.

Aos meus irmãos, Renata e Guilherme, pela torcida!

## RESUMO

SCHIAVO, Paula Fernanda. **Utilizando funções de (pré)-agregação derivadas da Integral de Choquet em Sistemas de Classificação Baseados em Regras *Fuzzy***. 2019. 55 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

O objetivo deste trabalho é propor o uso de funções de (pré)-agregação derivadas da integral de Choquet, para utilização em conjuntos de Sistemas de Classificação baseados em Regras *Fuzzy*, cuja tomada de decisão final pode ou não ser dada por funções de penalidade. Primeiramente, foi introduzido um método para criar medidas de confiança e suporte baseado em índices de *overlap*, que geralmente são usados para avaliar o grau de certeza ou interesse de uma determinada regra de associação. Estes índices de *overlap* são construídos a partir de funções de *overlap*, que são um tipo especial de funções de agregação, não necessariamente associativas, que servem para aplicações relacionadas aos problemas de sobreposição de conjuntos. Esta dissertação apresenta um novo Mecanismo de Raciocínio *Fuzzy* para ser usado em sistemas de classificação baseados em regras *fuzzy* considerando diferentes índices de *overlap*, que generaliza os métodos clássicos. Ao considerar vários índices de *overlap* e as funções de pré-agregação baseada na integral de Choquet para a tomada de decisão obtém-se a seleção da melhor classe, sem utilizar funções de penalidade. Por fim, é apresentado um exemplo detalhado de uma geração de conjuntos baseados em regras *fuzzy* e a seleção da melhor classe com base na abordagem proposta.

**Palavras-chave:** Sistemas Baseados em Regras *Fuzzy*, Raciocínio Aproximado, Integral Choquet, Funções de *Overlap*, Índices de *Overlap*, Funções de Pré-Agregação.

## ABSTRACT

SCHIAVO, Paula Fernanda. **Using (pre)-aggregation functions derived from the Integral of Choquet in Classification Systems Based on Fuzzy Rules.** 2019. 55 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

The purpose of this work is to offer the use of (pre) -aggregation functions derived from the Choquet integral, for use in Classification Systems based on Fuzzy Rules, whose final decision may or may not be given by penalty functions. Firstly, a method was introduced to create trust and support measures based on overlap indexes, which are generally used to evaluate the degree of certainty or interest of a given association rule. These overlap indices are built from overlap functions, which are a special type of aggregation functions, not necessarily associative, that serve for applications related to set overlapping problems. This dissertation presents a new Fuzzy Reasoning Mechanism to be used in classification systems based on fuzzy rules considering different indexes of overlap, which generalizes the classical methods. Considering several indexes of overlap and the pre-aggregation functions based on the Choquet integral for decision making, one obtains the best class selection without using penalty functions. Finally, a detailed example of a generation of sets based on fuzzy rules and the selection of the best class based on the proposed approach is presented.

**Keywords:** Systems Based on Fuzzy Rules, Choquet Integral, Overlap Functions, Overlap Indexes, Pre-Agregation Functions.

## LISTA DE FIGURAS

Figura 1	Termos Linguísticos para IMC e Idade . . . . .	37
Figura 2	Regras . . . . .	38

## LISTA DE TABELAS

Tabela 1	Exemplos de funções <i>overlap</i> . . . . .	18
Tabela 2	Exemplos de índices de <i>overlap</i> . . . . .	19
Tabela 3	Exemplos de índices de <i>overlap</i> construídos usando o Teorema 1 . . . . .	19
Tabela 4	Exemplos de funções de pré-agregação . . . . .	23
Tabela 5	Exemplos usados para treinamento . . . . .	37
Tabela 6	Graus de pertinência <i>fuzzy</i> . . . . .	39
Tabela 7	<i>Cnf</i> construído a partir de diferentes índices de <i>overlap</i> . . . . .	41
Tabela 8	Resultados do Algoritmo 2 . . . . .	44
Tabela 9	Resultados obtidos através do Algoritmo 3. Consenso . . . . .	44
Tabela 10	Resultados do Algoritmo 4 com Função Pré-Agregação $T_{HP}$ . . . . .	47
Tabela 11	Resultados do Algoritmo 4 com Função Pré-Agregação $TM$ . . . . .	47
Tabela 12	Resultados do Algoritmo 4 com Função Pré-Agregação $F_{NA}$ . . . . .	47
Tabela 13	Resultados do Algoritmo 4 com Função Pré-Agregação $F_{NA2}$ . . . . .	48
Tabela 14	Resultados do Algoritmo 4 com Função Pré-Agregação $C_{F_1F_2} GM - T_L$ . . . . .	48
Tabela 15	Resultados do Algoritmo 4 com Função Pré-Agregação $C_{F_1F_2} GM - F_{BPC}$ . . . . .	48

## LISTA DE ABREVIATURAS E SIGLAS

$CC$	Integral de Choquet Baseado em Cópula
$C_F$	Integral de Choquet Baseado em Função
$C_{F_1 F_2}$	Integral de Choquet Baseado em Duas Funções
CF	Conjunto <i>Fuzzy</i>
MRF	Mecanismo de Raciocínio <i>Fuzzy</i>
OWA	Ordered Weighted Averaging
SCBRF	Sistemas de Classificação Baseados em Regras <i>Fuzzy</i>

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	11
1.1	Objetivo Geral e Específicos	13
1.2	Organização do Texto	14
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	15
2.1	Conceitos Preliminares	15
2.2	Operadores de Agregação: T-Normas e Funções de <i>Overlap</i>	15
2.3	Índices de <i>Overlap</i>	17
2.4	Funções de Penalidade	19
2.5	Integral de Choquet	21
2.6	Funções de Pré-Agregação	22
2.6.1	Funções de Pré-Agregação Baseada na Integral de Choquet	23
2.6.2	Estado da Arte	26
<b>3</b>	<b>SCBRF USANDO VÁRIOS ÍNDICES DE <i>OVERLAP</i> E FUNÇÕES DE AGREGAÇÃO UTILIZANDO FUNÇÕES DE PENALIDADE PARA CONSENSO</b>	27
3.1	Sistemas de Classificação Baseados em Regras <i>Fuzzy</i>	27
3.2	Construção de Graus de Confiança de Índices de <i>Overlap</i>	28
3.3	Mecanismo de Raciocínio de um Sistema de Classificação Baseado em Regras <i>Fuzzy</i>	32
3.4	Usando Funções de Penalidade para Escolher a Melhor Classe em Conjuntos de Regras <i>Fuzzy</i>	34
3.5	Experimento	36
3.5.1	Avaliando os Graus de Confiança para Cada Regra	38
3.5.2	Inferência	41
3.6	Análise do Experimento	43
<b>4</b>	<b>APLICAÇÃO DAS FUNÇÕES DE (PRÉ)-AGREGAÇÃO BASEADAS NA INTEGRAL DE CHOQUET</b>	45
4.1	Metodologia	45
4.2	Alteração do Algoritmo 2	46
4.3	Experimento	46
4.4	Análise do Experimento	47
<b>5</b>	<b>CONCLUSÃO</b>	49
	REFERÊNCIAS	51

# 1 INTRODUÇÃO

Problemas relacionados à classificação estão presentes de diversas formas nas situações do cotidiano moderno. Os Sistemas de Classificação Baseados em Regras *Fuzzy* (SCBRF) são capazes de lidar com diversos problemas reais, conciliando modelos de descrição e tratamento de situações do mundo real e reais necessidades de representação. A essência dessa teoria está na aceitação da incerteza como um fato. Os conjuntos *fuzzy* permitem modelar a imprecisão e a incerteza que estão muitas vezes presentes em linguagens naturais. Por exemplo, o conjunto “pessoas altas” é *fuzzy*, pois, obviamente, existem diferentes graus de percepção de altura. Da mesma forma, os conjuntos de “números pequenos”, “pessoas acima do peso” e “dia frio” não possuem limites definidos, isto é, não existe um corte claro que discrimina objetos entre aqueles que estão no conjunto e aqueles que não estão. Um objeto pode pertencer simultaneamente a um conjunto *fuzzy* e seu complemento. Um conjunto *fuzzy*  $A$  definido em um conjunto de objetos  $X$  é representado por uma função de associação  $\mu_A : X \rightarrow [0, 1]$ , de tal forma que para qualquer objeto  $x \in X$  o valor  $\mu_A(x)$  mede o grau de associação de  $x$  no conjunto *fuzzy*  $A$  (BUSTINCE; RADKO 2001).

A teoria dos conjuntos *fuzzy* e os SCBRF têm sido extremamente úteis para resolver muitos problemas do mundo real, nos quais os dados são imprecisos como por exemplo: no processamento de imagens, segundo PAGOLA et al. (2012); em previsões, como demonstrado por RODRIGUES et al. (2016) em seu trabalho sobre previsões de pestes na agricultura; em problemas de classificação como desenvolvido por SANZ et al. (2015); e também dos controladores *fuzzy*, com ampla adoção, utilizados em sistemas industriais e eletrônicos de consumo, demonstrando que é notável o uso das aplicações práticas da lógica *fuzzy*.

Os sistemas baseados em regras, especialmente os SCBRFs, envolvem agregação de vários escores numéricos, que correspondem a graus de satisfação de antecedentes de regras (ZADEH, 1985). Uma regra pode ser uma declaração como

**SE  $x$  é  $A$  E  $y$  é  $B$  E  $z$  é  $C$  ENTÃO** alguma informação

Os antecedentes são geralmente valores de associação de  $x$  em  $A$ ,  $y$  em  $B$  e  $z$  em

$C$ . A força de “ativar” a regra é determinada por uma função de agregação que combina valores de associação  $f(\mu_A(x), \mu_B(y), \mu_C(z))$ .

A agregação é o processo de combinar vários valores de um mesmo tipo em um único valor do mesmo tipo. As funções matemáticas que fornecem um mecanismo para fazê-lo são chamadas de funções de agregação (BELIAKOV et al, 2016), que são funções monotônicas satisfazendo condições de fronteira. A teoria das funções de agregação tornou-se uma área de pesquisa em expansão nos últimos 30 anos. Os avanços teóricos são complementados por inúmeras aplicações em ciências da decisão, inteligência artificial, sistemas *fuzzy* e processamento de imagens. Uma das funções de agregação mais simples é a média aritmética, ou média, que produz um valor de saída em algum sentido representativo de suas entradas.

Uma função de agregação que está sendo muito utilizada em classificação (BARRENECHEA et al, 2013) e tomada de decisão (LOURENZUTTI et al, 2017) é a Integral de Choquet (CHOQUET, 1953). Esta função é definida em termos de uma medida *fuzzy*. Uma medida *fuzzy* permite atribuir importâncias a todos os subconjuntos possíveis e, portanto, oferece uma flexibilidade muito maior para modelar a incerteza (DENNEBERG, 1994; DUBOIS; PRADE, 2000; GRABISH, 2000; WANG; KLIR, 1992). Um subconjunto  $A \subseteq N$ , onde  $N = \{1, 2, \dots, n\}$  pode ser considerado como um grupo, de modo que uma medida fuzzy aplicada em  $A$  representa a importância ou o peso desse grupo. A condição de monotonicidade implica que adicionar novos elementos a um conjunto não diminui seu peso. (BELIAKOV et al, 2016)

A integral de Choquet generaliza a integral de Lebesgue e, o objetivo principal da agregação baseada na integral de Choquet (CHOQUET, 1953) é combinar as entradas de tal forma que não só a importância das entradas individuais (como nos significados ponderados), ou de sua magnitude (como nas funções médias ponderadas ordenadas), são levados em conta, mas também a importância de seus grupos (ou coligações) (BELIAKOV et al, 2016). Por exemplo, uma entrada específica pode não ser importante por si só, mas se torna muito importante na presença de algumas outras entradas. No diagnóstico médico, por exemplo, alguns sintomas por si só podem não ser realmente importantes, mas podem se tornar fatores-chave na presença de outros sinais.

Por outro lado, nos últimos tempos, há um crescente interesse em relaxar a condição de monotonicidade de modo que a classe de funções de agregação seja ampliada para abranger outras funções que são úteis do ponto de vista aplicado. Nesse sentido, surgem as funções de pré-agregação que são aquelas funções que satisfazem as mesmas condições de fronteira que as funções de agregação, mas que são monótonas somente ao longo de alguma direção fixa (LUCCA et al, 2016). A função moda é um exemplo de função de pré-agregação, mas não é considerado uma função de agregação justamente por não ser crescente em todo seu domínio.

Em (ELKANO et al, 2017), fase inicial desta dissertação, foi introduzido um novo

mecanismo de raciocínio *fuzzy* que leva em consideração as informações fornecidas por várias ou mesmo todas as regras *fuzzy* no sistema, através da aplicação de índices de *overlap*, que são usados para medir o grau de sobreposição entre dois conjuntos *fuzzy*, os resultados são produzidos por conjuntos de classificadores clássicos, com a combinação de diversas funções de *overlap* e índices de *overlap*, sendo que para obter um consenso desta classificação foi utilizado uma função de penalidade.

As funções de penalidade não produzem um novo valor de consenso, mas selecionam um dos resultados já obtidos, ou seja, aquele que apresenta o menor desvio em relação às demais alternativas. Como esta função torna a classificação custosa em processamento, propõe-se o uso das generalizações da Integral de Choquet como o operador de pré-agregação no mecanismo de raciocínio *fuzzy* e espera-se que as funções de pré-agregação classifiquem o exemplo proposto sem a necessidade de utilizar a função *penalty*.

## 1.1 Objetivo Geral e Específicos

O objetivo desta dissertação é criar um mecanismo de raciocínio *fuzzy* (MRF) que utilize funções de pré-agregação derivadas da Integral de Choquet em conjuntos de sistemas de classificação baseados em regras *fuzzy* através de índices de *overlap*, para seleção da melhor classe, sem a utilização de funções *penalty*, tendo como objetivo específicos:

1. Introduzir um método para construir medidas de confiança e suporte, com base em índices de *overlap* (Algoritmo 1);
2. Introduzir um novo MRF para o SCBRF, considerando diferentes índices de *overlap*, que generaliza os métodos clássicos (Algoritmo 2);
3. Gerar conjuntos baseados em regras *fuzzy*, considerando vários índices de *overlap* e funções de agregação, fornecendo resultados diferentes;
4. Desenvolver um método de consenso para a classificação, baseado em funções de penalidade (Algoritmo 3);
5. Apresentar resultados teóricos relacionados aos métodos desenvolvidos;
6. Desenvolver um exemplo envolvendo as etapas 1 a 4;
7. Alterar o Algoritmo 2 para que o MRF considere as funções de pré-agregação baseadas na integral de Choquet (Algoritmo 4);
8. Utilizar o mesmo exemplo para mostrar a não necessidade da função de penalidade na classificação ao executar as etapas 1 e 7.

A questão de pesquisa é baseada na ideia de que as funções de pré-agregação derivadas da integral de Choquet consideram uma melhor relação entre os elementos que estão sendo agregados, fornecendo a relevância de um conjunto, essa melhoria pode ser aplicada sobre a saída do grau de consistência (índice de *overlap*) do sistema para a classificação dos padrões para todas as classes.

Como contribuição científica demonstra-se a utilização de funções de pré-agregação derivadas da integral de Choquet (LUCCA et al, 2016) com índices de *overlap* em conjuntos de sistemas de classificação baseado em regras *fuzzy* para tomada de decisão em um experimento.

## 1.2 Organização do Texto

Esta dissertação esta organizada conforme mencionado a seguir.

O capítulo 2 apresenta a fundamentação teórica onde são mostrados os conceitos fundamentais utilizados no desenvolvimento desta proposta de dissertação. A seção 2.1, apresenta definições matemáticas relativas as funções e propriedades que serão citadas na continuidade da dissertação. Na Seção 2.2 são definidos os operadores de agregação, funções de agregação, t-normas e as funções de *overlap*. Na Seção 2.3 e 2.4, constam as definições e propriedades dos índices de *overlap* e funções de penalidade, respectivamente. A seção 2.5 apresenta algumas medidas *fuzzy* e a integral de Choquet. E na seção 2.6 constam as funções de pré-agregação baseada na integral de Choquet juntamente com uma visão do estado da arte dos artigos publicados até então.

Em seguida, o capítulo 3 apresenta o sistema de classificação baseado em regras *fuzzy* usando vários índices de *overlap* e funções de agregação utilizando funções de penalidade para consenso na tomada de decisão. A Seção 3.1 que retoma os sistemas de classificação baseados em regras *fuzzy*. A Seção 3.2 apresenta a construção de graus de confiança de índices de *overlap*. Na seção 3.3 é descrito o mecanismo de raciocínio de um sistema de classificação baseado em regras *fuzzy*. A seção 3.4 demonstra o uso de funções de penalidade para escolher a melhor classe em conjunto de regras *fuzzy*. Na seção 3.5 tem-se a aplicação de um experimento realizado através da teoria apresentada nas seções anteriores. E na seção 3.6 está descrita a análise dos resultados do experimento.

Na sequência, o capítulo 4 apresenta a aplicação das funções de (pré)-agregação baseadas na integral de Choquet no mecanismo de raciocínio *fuzzy*. A seção 4.1 descreve a metodologia utilizado na aplicação do mesmo experimento. A seção 4.2 apresenta a alteração do algoritmo desenvolvido para considerar as funções de (pré)-agregação. Nas seções 4.3 e 4.4, estão descritos a aplicação do experimento com o método proposto e a análise dos resultados, respectivamente. Por fim, o capítulo 5 sintetiza as considerações finais e os trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentadas os conceitos básicos que serão abordados no decorrer desta dissertação.

### 2.1 Conceitos Preliminares

Dado um conjunto de universo não-vazio  $U$ , denote por  $CF(U)$  o espaço de todos os conjuntos *fuzzy* definidos sobre  $U$ .

Um conjunto *fuzzy*  $A$  é dito normal se existir  $u \in U$  tal que  $X(u) = 1$ . Dois conjuntos *fuzzy*  $A, B \in CF(U)$  são ditos completamente disjuntos se  $A(u)B(u) = 0$ , para todo  $u \in U$ .

**Definição 1** Uma função  $F : [0, 1]^n \rightarrow \mathbb{R}$  é idempotente se,  $n > 1$  e para cada  $x \in [0, 1]$  tem-se que

$$F(x, \dots, x) = x.$$

### 2.2 Operadores de Agregação: T-Normas e Funções de *Overlap*

O conceito-chave no SCBRF são as funções de agregação:

**Definição 2** (BELIAKOV *et al*, 2007; MAYOR; TRILLAS, 1986) Uma função de agregação  $n$ -ária é um mapeamento  $M : [0, 1]^n \rightarrow [0, 1]$  satisfazendo as propriedades que seguem:

(A1)  $M$  é crescente<sup>1</sup> em cada argumento: para cada  $i \in \{1, \dots, n\}$ , se  $x_i \leq y$ , então

$$M(x_1, \dots, x_n) \leq M(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n);$$

(A2) As condições limites:  $M(0, \dots, 0) = 0$  e  $M(1, \dots, 1) = 1$ .

---

<sup>1</sup>Foi considerado que uma função crescente pode não ser estritamente crescente (e, analogamente, para as funções decrescentes).

**Definição 3** Uma função de agregação  $n$ -ária  $M : [0, 1]^n \rightarrow [0, 1]$  é dita  $0$ -positiva se também satisfizer a propriedade que segue:

(A3) As condições limites para  $0$ :  $M(x_1, \dots, x_n) = 0$  se e somente se  $x_i = 0$ , para todo  $i \in \{1, \dots, n\}$ .

Uma função de agregação  $n$ -ária  $M : [0, 1]^n \rightarrow [0, 1]$  é dita  $1$ -positiva se também satisfizer a propriedade que segue:

(A4) A condição limite  $1$ :  $M(x_1, \dots, x_n) = 1$  se e somente se  $x_i = 1$ , para todo  $i \in \{1, \dots, n\}$ .

**Definição 4** Uma função de agregação  $f : [0, 1]^n \rightarrow [0, 1]$  é dito que é do tipo média se for limitado pelo mínimo e máximo de seus argumentos, isto é, se para todos  $(x_1, \dots, x_n) \in [0, 1]^n$ , tem-se que:

$$\min\{x_1, \dots, x_n\} \leq f(x_1, \dots, x_n) \leq \max\{x_1, \dots, x_n\}.$$

Devido à monotonicidade das funções de agregação, o comportamento de médias é equivalente à propriedade de idempotência. (BELIAKOV et al, 2016)

**Definição 5** (LOWEN, 1996) Uma  $t$ -norma é uma função de agregação bivariada  $T : [0, 1]^2 \rightarrow [0, 1]$  satisfazendo as propriedades que seguem, para todo  $x, y, z \in [0, 1]$ :

(T1) Comutatividade:  $T(x, y) = T(y, x)$ ;

(T2) Associatividade:  $T(x, T(y, z)) = T(T(x, y), z)$ ;

(T3) Condições Limites:  $T(x, 1) = x$ .

Uma função de agregação  $T$  que satisfaz apenas (T3) e  $T(1, x) = x$  é chamada semi-copula.

Um elemento  $x \in ]0, 1]$  é um divisor zero não-trivial de  $T$  se existe  $y \in ]0, 1]$  tal que  $T(x, y) = 0$ . Uma  $t$ -norma é positiva se e somente se não tem divisores zero não triviais, isto é, se  $T(x, y) = 0$  então  $x = 0$  ou  $y = 0$ . Alguns exemplos de  $t$ -normas positivas e contínuas são

$$\begin{aligned} T_M(x, y) &= \min\{x, y\}, \\ T_P(x, y) &= xy. \end{aligned}$$

que são as  $t$ -normas do mínimo e do produto, respectivamente. Devido à propriedade associativa, trivialmente, pode-se definir  $t$ -normas  $n$ -ária.

**Definição 6** *BUSTINCE et al (2010)* Para um determinado vetor de pesos  $w = (w_1, \dots, w_n)$ ,  $w_i \geq 0$ ,  $\sum_{i=1}^n w_i = 1$ , a função OWA é dada por

$$OWA_w(\vec{x}) = \sum_{i=1}^n w_i x_{(i)} = (w, x_{\searrow})$$

As funções de média ponderada ordenada (do inglês, OWA) são funções de agregação simétrica que alocam pesos de acordo com o valor de entrada. Eles foram introduzidos por (YAGER, 1988). A notação  $x_{\searrow}$ , denota o vetor obtido de  $x$  que organiza seus elementos em ordem decrescente  $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)}$ .

**Definição 7** *BUSTINCE et al (2010)* Uma função overlap é uma função bivariada  $O: [0, 1]^2 \rightarrow [0, 1]$  satisfazendo as propriedades que seguem, para todo  $x, y \in [0, 1]$ :

- (O1)  $O$  é comutativa:  $O(x, y) = O(y, x)$ ;
- (O2)  $O(x, y) = 0$  se e somente se  $x = 0$  ou  $y = 0$ ;
- (O3)  $O(x, y) = 1$  se e somente se  $x = y = 1$ ;
- (O4)  $O$  é crescente;
- (O5)  $O$  é contínuo.

Uma função overlap  $O$  é associativa se e somente se  $O$  for uma t-norma contínua e positiva (veja BUSTINCE et al (2010)).

Exemplos de funções overlap estão presentes na Tabela 1 (conforme BUSTINCE et al (2010, 2012); JURIO et al (2013); BEDREGAL et al (2013); DIMURO; BEDREGAL (2014); DIMURO et al (2016); DIMURO; BEDREGAL (2015); DIMURO et al (2014,?); DIMURO; BEDREGAL (2015)). Observe que  $T_M$  e  $T_P$  também são t-normas.

**Proposição 1** (*GARCIA-JIMENEZ et al, 2015, Proposição 5*) Seja  $O: [0, 1]^2 \rightarrow [0, 1]$  uma função overlap e  $T: [0, 1]^n \rightarrow [0, 1]$  uma t-norma n-ária. Então, tem-se que

$$O(x, T(y_1, \dots, y_n)) = T(O(x, y_1), \dots, O(x, y_n))$$

se e somente se  $T = \min$ .

## 2.3 Índices de Overlap

Os índices de overlap são usados para medir o grau de sobreposição entre dois conjuntos fuzzy e consiste na generalização do índice de consistência de Zadeh entre dois conjuntos fuzzy sobre o mesmo universo referencial.

Tabela 1: Exemplos de funções *overlap*

Funções <i>Overlap</i>
$T_M(x, y) = \min\{x, y\}$
$T_P(x, y) = xy$
$O_p(x, y) = x^p y^p, p > 0$ (em particular, $O_{\sqrt{\cdot}}(x, y) = \sqrt{xy}$ )
$O_{DB}(x, y) = \begin{cases} \frac{2xy}{x+y} & \text{se } x + y \neq 0 \\ 0 & \text{se } x + y = 0 \end{cases}$
$O_2^V(x, y) = \begin{cases} \frac{1+(2x-1)^2(2y-1)^2}{2} & \text{se } x, y \in ]0.5, 1]; \\ \min\{x, y\} & \text{caso contrário.} \end{cases}$
$O_{m\frac{1}{2}}(x, y) = \min\{\sqrt{x}, \sqrt{y}\}$
$O_{mM}(x, y) = \min\{x, y\} \max\{x^2, y^2\}$
$O_k(x, y) = \min\{x^k y, x y^k\}$
$O_{rat}(x, y) = \frac{\sqrt{xy}}{\sqrt{xy}+1-xy}$

Várias definições de índices de *overlap* podem ser encontradas na literatura (ver, por exemplo: GARCIA-JIMENEZ et al (2015); BUSTINCE et al (2009); DUBOIS et al (2000); ZADEH (1978)). Neste artigo, adotamos a abordagem proposta por BUSTINCE et al (2009), que também foi formalizada por GARCIA-JIMENEZ et al (2015), juntamente com um método para construir índices de *overlap* por meio de funções de sobreposição. Nosso objetivo é definir a confiança e apoiar as medidas das regras de associação usando índices de *overlap*.

**Definição 8** (GARCIA-JIMENEZ et al, 2015) *Um índice de overlap é uma função  $\mathcal{O} : CF(U) \times CF(U) \rightarrow [0, 1]$  de modo que, para todo  $A, B, C \in CF(U)$ , as seguintes condições:*

- (O1)  $\mathcal{O}(A, B) = 0$  se e somente se  $A$  e  $B$  tem suporte disjuntivo, isto é, para todo  $u \in U$ , tem-se que  $A(u)B(u) = 0$ ;
- (O2)  $\mathcal{O}(A, B) = \mathcal{O}(B, A)$ ;
- (O3) Se  $B \leq C$ , então  $\mathcal{O}(A, B) \leq \mathcal{O}(A, C)$ .

*Um índice de overlap  $\mathcal{O}$  é dito normal sempre que a seguinte condição seja satisfeita:*

- (O4) Se existe  $u \in U$  tal que  $A(u) = B(u) = 1$ , então  $\mathcal{O}(A, B) = 1$ .

Alguns exemplos de índices de *overlap* estão presentes na Tabela 2. Observe que  $O_Z$  é o índice de consistência de ZADEH (1978).  $O_Z$  e  $O_\pi$  são normais e  $O_x$  não é normal para  $x \neq 1$ . (GARCIA-JIMENEZ et al, 2015)

Tabela 2: Exemplos de índices de *overlap*

Índices de <i>Overlap</i>
$\mathcal{O}_Z(A, B) = \max_{u \in U} \min\{A(u), B(u)\}$
$\mathcal{O}_x(A, B) = \begin{cases} 0 & \text{se } \forall u \in U : A(u)B(u) = 0 \\ x & \text{caso contrário,} \end{cases}$ para todo $x \in ]0, 1]$
$\mathcal{O}_\pi(A, B) = \frac{1}{n} \sum_{u \in U} A(u) \cdot B(u)$ , onde $n = \text{card}(U)$

Tabela 3: Exemplos de índices de *overlap* construídos usando o Teorema 1

Função de <i>Overlap</i> $O$	Função de Agregação $M$	Índice de <i>Overlap</i> $\mathcal{O}$
$O_\sqrt{\phantom{x}}$	média aritmética	$\mathcal{O}_\sqrt{\phantom{x}}(A, U) = \frac{1}{n} \sum_{u \in U} \sqrt{A(u) \cdot 1}$
$O_M$	máximo	$\mathcal{O}_Z(A, U) = \max_{u \in U} \min\{A(u), 1\}$ (ZADEH, 1978)
$O_{rat}$	média aritmética	$\mathcal{O}_{rat}(A, U) = \frac{1}{n} \sum_{u \in U} \frac{\sqrt{A(u) \cdot 1}}{\sqrt{A(u) \cdot 1 + 1 - A(u)}}$

**Teorema 1** *Seja  $M : [0, 1]^n \rightarrow [0, 1]$  uma função de agregação 0-positiva e  $O : [0, 1]^2 \rightarrow [0, 1]$  uma função de overlap. Então, o mapeamento  $\mathcal{O} : CF(U) \times CF(U) \rightarrow [0, 1]$  definido, para todo  $X, Y \in CF(U)$  e  $u_i \in U$ , com  $i = 1, \dots, n$ , como*

$$\mathcal{O}(X, Y) = M(O(X(u_1), Y(u_1)), \dots, O(X(u_n), Y(u_n))) \quad (1)$$

*é um índice de overlap (GARCIA-JIMENEZ et al, 2015). Em contra partida, se  $O$  é uma função de overlap e  $M : [0, 1]^n \rightarrow [0, 1]$  é uma função de agregação tal que  $\mathcal{O}$ , definida pela Equação (1), é um índice de overlap, então  $M$  é 0-positiva.*

A Tabela 3 mostra alguns exemplos de índices de *overlap* construídos usando o Teorema 1, usando algumas funções de sobreposição mostradas na Tabela 1.

## 2.4 Funções de Penalidade

Pode-se encontrar diferentes definições de funções de penalidade na literatura (ver, por exemplo YAGER (1993); YAGER; RYBALOV (1997); CALVO et al (2004); CALVO; BELIAKOV (2010); BELIAKOV (2014); WILKIN; BELIAKOV (2015)). Neste trabalho, considerando a discussão de Bustince et al. (BUSTINCE et al, 2016), decidimos adotar a seguinte definição, uma vez que supera muitas desvantagens das definições anteriores:

**Definição 9** (BUSTINCE et al, 2016) *Uma função  $f : [0, 1] \rightarrow \mathbb{R}$  é convexa se para todo  $x, y \in [0, 1]$  e para todo  $\lambda \in [0, 1]$ , tem-se a desigualdade  $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ .*

**Definição 10** (BUSTINCE et al, 2016) *Uma função  $f : [0, 1] \rightarrow \mathbb{R}$  é quasi-convexa se*

para todo  $x, y \in [0, 1]$  e para todo  $\lambda \in [0, 1]$  tem-se a desigualdade  $f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$ .

**Definição 11** (KURDILA; ZABARANKIN, 2005) Uma função  $f : [0, 1] \rightarrow \mathbb{R}$  é semi-contínua inferior em  $x_0 \in [0, 1]$  se

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0).$$

**Corolário 1** (KURDILA; ZABARANKIN, 2005, Corolário 2.1) Seja  $f : [0, 1] \rightarrow \mathbb{R}$  uma função semi-convexa e pouco semicontínua. Então, o conjunto de minimizadores de  $f$  é um conjunto não vazio conexo.

**Definição 12** Para qualquer intervalo fechado  $\mathbb{I} \subseteq \mathbb{R}$ , a função  $P : [0, 1]^{n+1} \rightarrow \mathbb{R}^+$  é uma função de penalidade se e somente se existir  $c \in \mathbb{R}^+$  tal que: (BUSTINCE et al, 2016, Definição 4.1)

(P1)  $P(\vec{x}, y) \geq c$ , para todo  $\vec{x} \in [0, 1]^n, y \in [0, 1]$ ;

(P2)  $P(\vec{x}, y) = c$  se e somente se  $x_i = y$ , para todo  $i = 1 \dots n$ , e

(P3)  $P$  é quasi-convexo e pouco semicontínuo em  $y$  para cada  $\vec{x} \in [0, 1]^n$ .

**Definição 13** (BUSTINCE et al, 2016, Definição 4.2) Seja  $P$  uma função de penalidade no sentido da Definição 12. A função  $f_P : [0, 1]^n \rightarrow [0, 1]$  é dita uma  $P$ -função, se, para cada  $\vec{x} \in [0, 1]^n$ , um tem que

$$f_P(\vec{x}) = \frac{a + b}{2}, \quad (2)$$

onde  $[a, b] = cl(Minz(P(\vec{x}, \cdot)))$ , e  $Minz(P(\vec{x}, \cdot))$  é o conjunto minimizador de  $P(\vec{x}, \cdot)$ , isso é,

$$Minz(P(\vec{x}, \cdot)) = \{y \in [0, 1] \mid P(\vec{x}, y) \leq P(\vec{x}, z), \text{ para cada } z \in [0, 1]\},$$

e  $cl(S)$  é o fecho de  $S \subseteq [0, 1]$ .

**Teorema 2** (BUSTINCE et al, 2016, Teorema 4.1) Uma função  $f : [0, 1]^n \rightarrow [0, 1]$  é uma  $P$ -função se e somente se  $f$  é idempotente.

Então, qualquer função de agregação do tipo média pode ser representada por uma  $P$ -função.

**Exemplo 1** Considerando uma função idempotente  $f : [0, 1]^n \rightarrow [0, 1]$ ,  $\epsilon > 0$  e  $c \geq 0$ , a função  $P_f : [0, 1]^{n+1} \rightarrow \mathbb{R}^+$ , definida, para todo  $\vec{x} \in [0, 1]^n$  e  $y \in [0, 1]$  então:

$$P_f(\vec{x}, y) = \begin{cases} c & \text{se } x_i = y \text{ para cada } i \\ |f(\vec{x}) - y| + c + \epsilon & \text{caso contrário} \end{cases} \quad (3)$$

é uma função de penalidade não contínua (BUSTINCE et al, 2016, Prova do Teorema 4.1). Agora, definindo a função  $P'_f : [0, 1]^{n+1} \rightarrow \mathbb{R}^+$ , para todo  $\vec{x} \in [0, 1]^n$  e  $y \in [0, 1]$ , então:

$$P'_f(\vec{x}, y) = |f(\vec{x}) - y| + V(\vec{x}) + c, \quad (4)$$

onde  $V$  é uma medida rígida de propagação contínua (GAGOLEWSKI, 2015). Segue que  $P_f$  é uma função de penalidade contínua (BUSTINCE et al, 2016, Exemplo 4.1).

A função de penalidade  $P$  descreve a dissimilaridade ou desacordo entre uma entrada arbitrária  $\vec{x}$  e um valor  $y$ . Então, a  $P$ -função  $f$  é uma função que minimiza a dissimilaridade escolhida.

## 2.5 Integral de Choquet

**Definição 14** (CHOQUET, 1953; SUGENO, 1974) Seja  $N = \{1, \dots, n\}$ . Uma medida fuzzy é uma função  $m : 2^N \rightarrow [0, 1]$  que é monotônica (ou seja,  $m(A) \leq m(B)$  sempre que  $A \subset B$ ) e satisfaz  $m(\emptyset) = 0$  e  $m(N) = 1$ .

No contexto das funções de agregação, medidas fuzzy são usadas para modelar a importância de um conjunto, ou seja, a relação entre os elementos a serem agregados.

Exemplos de medidas fuzzy:

(1) Medida Uniform  $m_U(A) = \frac{|A|}{n}$

(2) Medida Dirac: Seja  $i \in N$  fixo,  $m_D^i(A) = \begin{cases} 1 & \text{se } i \in A \\ 0 & \text{se } i \notin A \end{cases}$

(3) Medida Aditiva (Wmean): Seja  $W = (w_1, \dots, w_n) \in [0, 1]^n$  tal que  $\sum_{i=1}^n w_i = 1$ .  
Define

$$m_W(\{i\}) = w_i$$

Então, para  $|A| \geq 1$ :

$$m_W(A) = \sum_{i \in A} w_i$$

(4) Medida Simétrica (OWA): Seja  $W = (w_1, \dots, w_n) \in [0, 1]^n$  tal que  $\sum_{i=1}^n w_i = 1$ .  
Então, para cada subconjunto não vazio de  $A$ , definimos:

$$m_{sW}(A) = \sum_{i=1}^{|A|} w_i$$

(5) Medida de Poder

$$m_{PM}(A) = \left(\frac{|A|}{n}\right)^q$$

com  $q > 0$

**Definição 15** (CHOQUET, 1953) *Seja  $m$  uma medida fuzzy em  $N$ . Dizemos que  $m$  é:*

- *Aditivo se para qualquer subconjunto disjunto  $A, B \subseteq N, m(A \cup B) = m(A) + m(B)$ ;*
- *Simétrico se para qualquer subconjunto  $A, B \subseteq N, |A| = |B|$  implica  $m(A) = m(B)$ ;*

**Definição 16** (CHOQUET, 1953) *Seja  $m$  uma medida fuzzy em  $N$  e  $x \in [0, \infty]^n$ . A integral discreta de Choquet de  $x$  em relação a  $m$  é definida por:*

$$\mathfrak{C}_m(\vec{x}) = \sum_{i=1}^n (x_{(i)} - x_{(i-1)}) \cdot m(A_{(i)}) \quad (5)$$

onde  $(.)$  é uma permutação de  $\{1, \dots, n\}$  tal que  $0 \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  com a convenção  $x_{(0)} = 0$  e  $A_{(i)} = \{i, \dots, n\}$

Observe que a Equação 5 também pode ser escrita como:

$$\mathfrak{C}_m(\vec{x}) = \sum_{i=1}^n (x_{(i)} \cdot m(A_{(i)}) - x_{(i-1)} \cdot m(A_{(i)})) \quad (6)$$

## 2.6 Funções de Pré-Agregação

Em (LUCCA et al, 2016) é definida as funções de pré-agregação como aquelas funções que satisfazem as mesmas condições de fronteira que as funções de agregação, mas que são monótonas ao longo de alguma direção fixa. Ou seja, as funções de pré-agregação são apenas funções direcionalmente monótonas que satisfazem condições de contorno apropriadas. A relevância dessa noção surge, por um lado, do fato de que ela permite generalizar noção de função de agregação e, por outro lado, de sua aplicabilidade.

Vários trabalhos foram propostos para aplicar funções de agregação e de pré-agregação (com características de média e idempotência) para agregar as informações locais associadas a cada regra. A ideia inicial foi proposta por (BARRENECHEA et al, 2013), onde a integral Choquet foi utilizada para realizar essa agregação de uma forma que também levou em conta a correlação entre as regras. Depois disso, esse método foi aprimorado por (LUCCA et al, 2016), introduzindo uma generalização da integral de Choquet onde o operador de produto desta função é substituído por uma t-norma, resultando em uma pré-agregação que não é uma função de agregação. Em (LUCCA et al, 2016), a integral de Choquet em sua forma expandida foi generalizada usando funções de cópula (ALSINA; SCHWEIZER, 2006), ao invés do operador de produto, obtendo funções de agregação chamadas CC-integrais.

Tabela 4: Exemplos de funções de pré-agregação

Pré-agregação	Definição	Família
Mínimo	$T_M(x, y) = \min\{xy\}$	t-norma, cópula
Łukasiewicz	$T_{\mathbb{L}}(x, y) = \max\{0, x + y - 1\}$	t-norma, cópula
Produto Hamacher	$T_{HP}(x, y) = \begin{cases} 0 & \text{se } x = y = 0 \\ \frac{xy}{x+y-xy} & \text{caso contrário} \end{cases}$	t-norma, cópula
Média geométrica	$GM(x, y) = \sqrt{xy}$	função overlap
$F_{BPC}$	$F_{BPC}(x, y) = xy^2$	função agregação
$F_{NA}$	$F_{NA}(x, y) = \begin{cases} x & \text{se } x \leq y \\ \min\{\frac{x}{2}, y\} & \text{caso contrário} \end{cases}$	função de pré-agregação
$F_{NA2}$	$F_{NA2}(x, y) = \begin{cases} 0 & \text{se } x = 0 \\ \frac{x+y}{2} & \text{se } 0 < x \leq y \\ \min\{\frac{x}{2}, y\} & \text{caso contrário} \end{cases}$	função de pré-agregação

**Definição 17** (LUCCA et al, 2016) Uma função de pré-agregação é uma função  $F : [0, 1]^n \rightarrow [0, 1]$  tal que:

(P1)  $F$  é  $\vec{r}$ -crescente para algum vetor real  $\vec{r} \in [0, 1]^n$  ( $\vec{r} \neq \vec{0}$ );

(P2)  $F(0, \dots, 0) = 0$  e  $F(1, \dots, 1) = 1$ ;

Observe que se  $F : [0, 1]^n \rightarrow [0, 1]$  é uma função de agregação, então  $F$  é também uma função de pré-agregação, desde que, por exemplo,  $(1, 0, \dots, 0)$ -crescente.

A Tabela 4 mostra alguns exemplos de funções de pré agregação conforme (LUCCA et al, 2016). Nesta tabela as funções  $T_M$ ,  $T_{\mathbb{L}}$  e  $T_{HP}$  são funções de agregação, denominadas t-normas.

### 2.6.1 Funções de Pré-Agregação Baseada na Integral de Choquet

**Definição 18** (LUCCA et al, 2016) Seja  $m : 2^N \rightarrow [0, 1]$  uma medida fuzzy e  $M : [0, 1]^2 \rightarrow [0, 1]$  uma função tal que  $M(0, x) = 0$  para  $x \in [0, 1]$ . Tomando como base a integral Choquet, definimos a função  $\mathfrak{C}_m^M : [0, 1]^n \rightarrow [0, n]$ , segue

$$\mathfrak{C}_m^M(\vec{x}) = \sum_{i=1}^n M(x_{(i)} - x_{(i-1)}, m(A_{(i)})) \quad (7)$$

onde  $N = \{1, \dots, n\}$ ,  $(x_{(1)}, \dots, x_{(n)})$  é uma permutação crescente de  $x$ , isso é  $0 \leq x_{(1)} \leq \dots \leq x_{(n)}$ , com a convenção que  $x_{(0)} = 0$ , e  $A_{(i)} = \{(i), \dots, (n)\}$  é o subconjunto de índices de  $n - i + 1$  maiores componentes de  $\vec{x}$ .

É imediato que toda t-norma  $T$  satisfaz as condições da Definição 18. Em (LUCCA et al, 2016) foi provado que para toda t-norma  $T$  tem-se que  $\mathfrak{C}_m^T$  é  $(1, \dots, 1)$ -pré-agregação.

**Definição 19** Uma função bivariada  $C : [0, 1]^2 \rightarrow [0, 1]$  é dito ser uma cópula, se para todo  $x, x', y, y' \in [0, 1]$  com  $x \leq x'$  e  $y \leq y'$ , nas seguintes condições:

$$(C1) \quad C(x, y) + C(x', y') \geq C(x, y') + C(x', y);$$

$$(C2) \quad C(x, 0) = C(0, x) = 0;$$

$$(C3) \quad C(x, 1) = C(1, x) = 1;$$

As cópulas são funções que ligam as funções de distribuição de probabilidade (bidimensional) às suas margens unidimensionais, desempenhando um papel importante na teoria dos espaços métricos probabilísticos e na estatística. (ALSINA; SCHWEIZER, 2006)

**Definição 20** (LUCCA et al, 2016) Seja  $m : 2^N \rightarrow [0, 1]$  uma medida fuzzy e  $C : [0, 1]^2 \rightarrow [0, 1]$  uma cópula bivariada. A integral de Choquet baseado em cópula (CC-integral) em relação a  $m$  é definido como uma função  $\mathfrak{C}_m^C : [0, 1]^n \rightarrow [0, 1]$ , para todo  $x \in [0, 1]^n$ , segue

$$\mathfrak{C}_m^C(\vec{x}) = \sum_{i=1}^n C(x_{(i)}, m(A_{(i)})) - C(x_{(i-1)}, m(A_{(i)})) \quad (8)$$

onde  $(x_{(1)}, \dots, x_{(n)})$  é uma permutação crescente de  $\vec{x}$ , isso é  $0 \leq x_{(1)} \leq \dots \leq x_{(n)}$ , com a convenção que  $x_{(0)} = 0$ , e  $A_{(i)} = \{(i), \dots, (n)\}$  é o subconjunto de índices de  $n - i + 1$  maiores componentes de  $\vec{x}$

**Teorema 3** (LUCCA et al, 2016, Teorema 4) Para qualquer cópula bivariada  $C : [0, 1]^2 \rightarrow [0, 1]$  e uma medida fuzzy  $m : 2^N \rightarrow [0, 1]$ ,  $\mathfrak{C}_m^C$  é uma função de agregação média.

**Definição 21** (LUCCA et al, 2018) Seja  $F : [0, 1]^2 \rightarrow [0, 1]$  uma função bivariada e  $m : 2^N \rightarrow [0, 1]$  uma medida fuzzy. A integral de Choquet baseada em  $F$  ( $C_F$ -integral) em relação a  $m$  é a função  $\mathfrak{C}_m^F : [0, 1]^n \rightarrow [0, 1]$ , para todo  $x \in [0, 1]^n$ , segue

$$\mathfrak{C}_m^F(\vec{x}) = \min \left\{ 1, \sum_{i=1}^n F(x_{(i)} - x_{(i-1)}, m(A_{(i)})) \right\} \quad (9)$$

onde  $(x_{(1)}, \dots, x_{(n)})$  é uma permutação crescente de  $\vec{x}$ , isso é  $0 \leq x_{(1)} \leq \dots \leq x_{(n)}$ , com a convenção que  $x_{(0)} = 0$ , e  $A_{(i)} = \{(i), \dots, (n)\}$  é o subconjunto de índices de  $n - i + 1$  maiores componentes de  $\vec{x}$

**Definição 22** (LUCCA et al, 2018) Uma função bivariada  $F : [0, 1]^2 \rightarrow [0, 1]$  com 0 como elemento aniquilador à esquerda, isto é, satisfazendo:

$$(LAE) \quad \forall y \in [0, 1] : F(0, y) = 0,$$

é dito ser 0-absorvente à esquerda.

Além disso, as duas propriedades básicas a seguir também são importantes:

**(RNE)** Elemento Neutro à Direita:  $\forall x \in [0, 1] : F(x, 1) = x$ ;

**(LC)** Propriedade Conjunta à Esquerda:  $\forall x, y \in [0, 1] : F(x, y) \leq x$ ;

Qualquer função bivariada  $F : [0, 1]^2 \rightarrow [0, 1]$  satisfazendo ambos **(LAE)** e **(RNE)** é chamada função **(RNE)** 0-absorvente à esquerda

**Teorema 4** (LUCCA et al, 2018, Teorema 1) Para qualquer medida fuzzy  $m : 2^N \rightarrow [0, 1]$  e **(RNE)** função 0-absorvente à esquerda  $F : [0, 1]^2 \rightarrow [0, 1]$ ,  $\mathfrak{C}_m^F$  é uma função de  $\vec{1}$ -pré-agregação.

**Teorema 5** (LUCCA et al, 2018, Teorema 2) Para qualquer medida fuzzy  $m : 2^N \rightarrow [0, 1]$  e 0-absorvente à esquerda função  $(1, 0)$ -pré-agregação  $F : [0, 1]^2 \rightarrow [0, 1]$ ,  $\mathfrak{C}_m^F$  é uma função de 1-pré-agregação.

(LUCCA et al, 2018) Um conceito importante usado nas  $C_{F_1 F_2}$ -integral é a propriedade de dominância (ou, inversamente, subordinação):

**(DM)**  $F_1$ -Dominância (ou, equivalente,  $F_2$ -Subordinação):  $F_1 \geq F_2$ , isso é:  $\forall x, y \in [0, 1] : F_1(x, y) \geq F_2(x, y)$

**Definição 23** (LUCCA et al, 2018) Seja  $m : 2^N \rightarrow [0, 1]$  uma medida fuzzy simétrica e  $F_1, F_2 : [0, 1]^2 \rightarrow [0, 1]$  duas funções de fusão cumprindo:

(i1)  $F_1$  - dominância;

(i2)  $F_1$  é crescente.

Uma  $C_{F_1 F_2}$ -integral é definida como uma função  $\mathfrak{C}_m^{F_1 F_2} : [0, 1]^n \rightarrow [0, 1]$ , para todo  $x \in [0, 1]^n$ , segue

$$\mathfrak{C}_m^{F_1 F_2}(\vec{x}) = \min \left\{ 1, x_{(1)} + \sum_{i=2}^n F_1(x_{(i)}, m(A_{(i)})) - F_2(x_{(i-1)}, m(A_{(i)})) \right\} \quad (10)$$

onde  $(x_{(1)}, \dots, x_{(n)})$  é uma permutação crescente de  $\vec{x}$ , isso é  $0 \leq x_{(1)} \leq \dots \leq x_{(n)}$ , com a convenção que  $x_{(0)} = 0$ , e  $A_{(i)} = \{(i), \dots, (n)\}$  é o subconjunto de índices de  $n - i + 1$  maiores componentes de  $\vec{x}$

## 2.6.2 Estado da Arte

Em (LUCCA et al, 2016) são apresentados as definições, propriedades e construção das funções de pré-agregação em raciocínio *fuzzy*.

No trabalho (LUCCA et al, 2016) é apresentado o conceito de função de agregação baseada na integral de Choquet usando diferentes Copulas, incluindo t-normas, funções de *overlap* e copulas que não são t-normas e também não são *overlaps*. Foram aplicados em 30 base de dados e apresentado um teste estatístico da análise da aplicação da CC-integrais em problemas de classificação. Mostrando que a CC-integral baseado na t-norma do mínimo apresentou os melhores resultados entre as CC-integrais consideradas no trabalho.

Em (DIMURO et al, 2017) foi estudado o comportamento da CC-integral em SC-BRF, quando se considera a t-norma do mínimo como copula, a é denominada de CMin-integral. Apresentado que a CMin-integral obtém um desempenho que é, com um alto nível de confiança, melhor que a abordagem que adota a regra vencedora (máximo) no MRF.

No artigo (LUCCA et al, 2018) é apresentado uma família de  $C_F$  Integral, generalizando a integral de Choquet e substituindo o operador do produto por uma função  $F$  com algumas propriedades fracas. É apresentado um teste estatístico em 33 base de dados públicas, que comprovam que a proposta de  $C_F$ -integrais não-média obtêm resultados mais precisos que as  $C_F$ -integrais médias, oferecendo assim novas possibilidades para agregar com precisão as informações no MRF.

O artigo (LUCCA et al, 2018) apresenta a família de funções  $C_{F_1F_2}$ -integral, que é uma generalização da integral de Choquet baseada na CC-integral, com base nos pseudo-pares de funções de pré-agregação  $(F_1, F_2)$ . As funções candidatas a serem  $F_1$ , possuem características de dominância, e as funções candidatas ao  $F_2$ , possuem característica de subordinação. Foi comparado o desempenho da  $C_{(F_1F_2)}$ -integral que atingiu a maior precisão com os pares estudados no artigo, contra a melhor função não-média do trabalho anterior  $C_F$ -integral, obtendo resultados estatisticamente melhores do que MRF clássicas da combinação aditiva e soma probabilística

### 3 SCBRF USANDO VÁRIOS ÍNDICES DE *OVERLAP* E FUNÇÕES DE AGREGAÇÃO UTILIZANDO FUNÇÕES DE PENALIDADE PARA CONSENSO

Este capítulo apresenta o sistema de classificação baseado em regras *fuzzy* usando vários índices de *overlap* e funções de agregação, juntamente com a construção de graus de confiança de índices de *overlap*. Descrevemos o mecanismo de raciocínio de um sistema de classificação baseado em regras *fuzzy* e também a escolha da melhor classe em conjuntos de regras *fuzzy* usando funções de penalidade. Ao final consta a aplicação da teoria estudada em um experimento juntamente com a avaliação dos resultados.

#### 3.1 Sistemas de Classificação Baseados em Regras *Fuzzy*

Um problema de classificação, do ponto de vista da aprendizagem supervisionada, consiste em encontrar uma regra de decisão que permita determinar a classe de um novo objeto (também chamado de exemplo  $\chi \in E$ ) entre as classes já existentes e conhecidas,  $C \in \{C_1, \dots, C_M\}$ . Um exemplo é descrito através de um conjunto de observações  $\chi = (\chi_1, \dots, \chi_n)$ . Cada uma dessas observações é chamada de variável, atributo ou característica.

Dentre todas as técnicas utilizadas para enfrentar problemas de classificação, um dos mais utilizados são os SCBRF, pois permitem a inclusão de todas as informações disponíveis na modelagem do sistema, gerando um modelo interpretável e fornecendo resultados precisos. Os dois componentes principais de SCBRFs são:

- A Base de Conhecimento contendo a Base de Regras e a Base de Dados. No primeiro, as regras de inferência *fuzzy* são armazenadas. O segundo inclui as funções de associação.
- O Mecanismo de Raciocínio *Fuzzy*, que faz uso das informações na Base de Conhecimento para classificar os exemplos

O design de um classificador ou um sistema de classificação pode ser visto como a

busca de um mapeamento:

$$D : E \rightarrow \{C_1, \dots, C_M\}$$

otimizado no sentido de algum critério que determina a qualidade do classificador.

Um classificador *fuzzy* é um sistema baseado em regras *fuzzy* que utiliza a incerteza apenas no mecanismo de raciocínio (RIID; RUSTERN, 2014) e que consiste em regras no seguinte formato:

Regra  $R_q$ : Se  $\chi_1$  é  $A_{q1}$  e  $\dots$  e  $\chi_n$  é  $A_{qn}$  então Classe  $C_q \in \{C_1, \dots, C_M\}$  com  $Cnf_q$

onde  $R_q$  é o rótulo da regra,  $\chi = (\chi_1, \dots, \chi_n)$  é um vetor n-dimensional de exemplo,  $A_{qi}$  denota o rótulo linguístico do i-ésimo recurso associado à regra  $q$ ,  $C_q$  é um consequente da classe e  $Cnf_q \in [0, 1]$  é o grau de confiança da regra  $R_q$  (ou seja, o peso da regra).

### 3.2 Construção de Graus de Confiança de Índices de *Overlap*

Para calcular o grau de confiança da regra  $R_q$ , ou seja, para calcular  $Cnf_q$ , pode-se usar o grau de confiança e o suporte. A confiança de uma associação é classicamente medida pela co-ocorrência de atributos em tuplas no banco de dados.

Considere um conjunto de  $p$  regras  $\{R_1, \dots, R_p\}$  conforme segue:

Regra  $R_1$ : Se  $\chi_1$  é  $A_{11}$  e  $\dots$  e  $\chi_n$  é  $A_{1n}$  então Classe  $C_1$  com  $Cnf_1 = ?$

...

Regra  $R_p$ : Se  $\chi_1$  é  $A_{p1}$  e  $\dots$  e  $\chi_n$  é  $A_{pn}$  então Classe  $C_p$  com  $Cnf_p = ?$

e pegue um conjunto de  $m$  exemplos  $\chi_l$  com  $l = 1, \dots, m$ .

A seguir, apresentamos o Algoritmo 1, que foi desenvolvido para criar medidas de confiança e suporte, usando índices de *overlap*, no treinamento dos elementos que já estão classificados. Este algoritmo é uma generalização do comumente usado em (RIID; RUSTERN, 2014).

---

**Algoritmo 1**


---

**Entrada:** Um conjunto de regras  $R_j$ , com  $j \in \{1, \dots, p\}$ , e um conjunto de exemplos  $\chi_l$ , com  $l \in \{1, \dots, m\}$ .

**Saída:** As medidas de Confiança ( $Cnf(R_j)$ ) e Suporte ( $Supp(R_j)$ ), para cada regra  $R_j$ , com  $j \in \{1, \dots, p\}$ .

- 1: Selecionar uma função de *overlap*  $O$  e um índice de *overlap*  $\mathcal{O}$ ;
- 2: Ordenar os exemplos  $\chi_l = (\chi_{l1}, \dots, \chi_{ln})$ , com  $l = \{1, \dots, m\}$ , tendo em conta a classe que classificam;
- 3: **for**  $q = 1$  to  $p$  **do**
- 4: Selecionar os  $s \leq m$  exemplos que nos dizem que o objeto considerado pertence à classe  $C_q$  associada à regra  $R_q$ ;
- 5: **for**  $j = 1$  to  $s$  **do**
- 6: Calcular (Grau de pertinência)

$$c_j(\chi_j) = O(A_{q1}(\chi_{j1}), \dots, A_{qn}(\chi_{jn})); \quad (11)$$

- 7: **end for**
- 8: Construir o conjunto *fuzzy* em  $U$

$$\mathbb{C}_{qs} = \{(u_1, c_1(\chi_1)), \dots, (u_s, c_s(\chi_s)), (u_{s+1}, 0), \dots, (u_m, 0)\}; \quad (12)$$

- 9: **for**  $l = 1$  to  $m$  **do**
- 10: Calcular

$$c_l(\chi_l) = O(A_{q1}(\chi_{l1}), \dots, A_{qn}(\chi_{ln})); \quad (13)$$

- 11: **end for**
- 12: Construir o conjunto *fuzzy* em  $U$

$$\mathbb{C}_{qm} = \{(u_1, c_1(\chi_1)), \dots, (u_m, c_m(\chi_m))\};$$

- 13: Calcular

$$Cnf(R_q) = \frac{\mathcal{O}(\mathbb{C}_{qs}, U)}{\mathcal{O}(\mathbb{C}_{qm}, U)}; \quad (14)$$

- 14: Calcular

$$Supp(R_q) = \mathcal{O}(\mathbb{C}_{qs}, U); \quad (15)$$

- 15: **end for**
- 

**Observação 1** Observe que:

1. Se, na Equação (14) e (15), nós tomamos a expressão para o índice de overlap  $\mathcal{O}_\pi$ ,

definido na Tabela 2, recuperamos o grau de confiança e o suporte apresentado por Ishibuchi et al. (ISHIBUCHI et al, 2005), que são dadas, respectivamente, por:

$$\begin{aligned}
 Cnf(R_q) &= \frac{\mathcal{O}_\pi(\mathbb{C}_{qs}, U)}{\mathcal{O}_\pi(\mathbb{C}_{qm}, U)} \\
 &= \frac{\frac{1}{m} \sum_{i=1}^s c_i(\chi_i)}{\frac{1}{m} \sum_{i=1}^m c_i(\chi_i)} \\
 &= \frac{\sum_{i=1}^s c_i(\chi_i)}{\sum_{i=1}^m c_i(\chi_i)} \tag{16}
 \end{aligned}$$

e

$$Supp(R_q) = \mathcal{O}_\pi(\mathbb{C}_{qs}, U) = \frac{1}{m} \sum_{i=1}^s c_i(\chi_i). \tag{17}$$

2. Na construção do grau de confiança  $Cnf$  da regra, usando a Equação (14), só levamos em consideração os conjuntos  $\mathbb{C}_{qs}$  e  $\mathbb{C}_{qm}$ . Existem situações em que é necessário considerar o índice de overlap do conjunto composto pelas regras que não classificam a classe considerada nesse momento (SANZ et al, 2010). No Algoritmo 1, isso significa que devemos construir o conjunto

$$\mathbb{C}_{q(m-s)} = \{(u_1, 0), \dots, (u_s, 0), (u_{s+1}, 0), (u_{s+1}, c_{s+1}(\chi_{s+1})), \dots, (u_m, c_m(\chi_m))\}$$

e construir o grau de confiança da seguinte forma:

$$Cnf(R_q) = \frac{\max\{0, \mathcal{O}(\mathbb{C}_{qs}, U) - \mathcal{O}(\mathbb{C}_{q(m-s)}, U)\}}{\mathcal{O}(\mathbb{C}_{qm}, U)}. \tag{18}$$

**Teorema 6** Seja  $M : [0, 1]^2 \rightarrow [0, 1]$  uma função de agregação que, para todo  $x, y \in [0, 1]$ , tal que

$$M(x, y) = f^{-1}(\omega f(x) + (1 - \omega)f(y)),$$

onde  $\omega \in ]0, 1[$  e  $f : [0, 1] \rightarrow [a, b]$  é uma função contínua e estritamente crescente. Então, tem-se:

(i)  $M(x, \min\{y_1, \dots, y_n\}) = \min\{M(x, y_1), \dots, M(x, y_n)\}$ ;

(ii)  $M$  é uma função de agregação 0-positiva.

Demonstração. Segue que:

(i) Considere  $L = \min\{y_1, \dots, y_n\}$ . Então tem-se que  $M(x, L) \leq M(x, y_i)$ , para todo  $i = 1, \dots, n$ , e

$$\min\{M(x, y_1), \dots, M(x, y_n)\} = M(x, L) = M(x, \min\{y_1, \dots, y_n\}).$$

(ii) Existe um  $M(x, y) = 0$  se e somente se  $f^{-1}(\omega f(x) + (1 - \omega)f(y)) = 0$  se e somente se  $\omega f(x) + (1 - \omega)f(y) = f(0)$ . Se  $x$  ou  $y$  são estritamente positivos, então, da monotonicidade de  $f$ , temos que  $f(x)$  ou  $f(y)$  é estritamente maior que  $f(0)$ , e então  $\omega f(x) + (1 - \omega)f(y) > 0$ . Sendo que  $x = y = 0$ .

**Corolário 2** No Teorema 6, tem-se a seguinte desigualdade:

$$M(\min\{x_1, \dots, x_n\}, \min\{y_1, \dots, y_n\}) \leq \min\{M(x_1, y_1), \dots, M(x_n, y_n)\}. \quad (19)$$

PROVA. Isso decorre da monotonicidade de  $M$ .

Usando a notação do Passo 6 do Algoritmo 1, criamos os seguintes conjuntos sobre o referencial  $U$ :

$$\binom{s}{A_{qi}}_{\gamma} = \{(u_l, A_{qi}(\chi_{l\gamma})) | u_l \in U\}, \quad (20)$$

onde  $\gamma = 1, \dots, n$ ,  $q \in \{1, \dots, p\}$  é o rótulo da regra  $R_q$ ,  $s \leq m$  é o número de exemplos que informa que o objeto considerado pertence à classe  $C_q$  associada à regra  $R_q$ ,  $m$  é o número total de exemplos e  $\chi_l = (\chi_{l1}, \dots, \chi_{ln})$ , com  $l = \{1, \dots, m\}$ , é o conjunto ordenado de exemplos, levando em consideração a classe que classificam.

Então, apresentamos o seguinte resultado:

**Teorema 7** Seja  $O : [0, 1]^2 \rightarrow [0, 1]$  uma função overlap e  $M : [0, 1]^n \rightarrow [0, 1]$  uma função de agregação 0-positiva satisfazendo as condições do Teorema 6. Seja  $\mathcal{O} : CF(U) \times CF(U) \rightarrow [0, 1]$  um índice de overlap construído, para  $O$  e  $M$ , de acordo com o Teorema 1. Então na configuração do Algoritmo 1, sempre que se considere  $T = \min$ , tem-se que:

$$\mathcal{O}(C_{qs}, U) \leq \min \left\{ \mathcal{O} \left( \binom{s}{A_{qi}}_1, U \right), \dots, \mathcal{O} \left( \binom{s}{A_{qi}}_n, U \right) \right\},$$

onde  $C_{qs}$  é o conjunto fuzzy em  $U$  definido pela Equação (12),  $q \in \{1, \dots, p\}$  é o rótulo da regra  $R_q$ ,  $s \leq m$  é o número de exemplos que informa que o objeto considerado pertence à classe  $C_q$  associado para a regra  $R_q$ , e  $m$  é o total do número de exemplos.

PROVA. Isso segue que:

$$\begin{aligned}
& \mathcal{O}(\mathbb{C}_{qs}, U) \\
&= \overset{s}{M} \left( \mathcal{O}(\min\{A_{q1}(\chi_{i1}), \dots, A_{qn}(\chi_{in})\}, 1) \right) \text{ pelas Equações (1) e (11)} \\
&= M \left( \mathcal{O}(\min\{A_{q1}(\chi_{11}), \dots, A_{qn}(\chi_{1n})\}, 1), \dots, \right. \\
&\quad \left. \mathcal{O}(\min\{A_{q1}(\chi_{s1}), \dots, A_{qn}(\chi_{sn})\}, 1) \right) \\
&= M \left( \min \left\{ \mathcal{O}(A_{q1}(\chi_{11}), 1), \dots, \mathcal{O}(A_{qn}(\chi_{1n}), 1) \right\}, \dots, \right. \\
&\quad \left. \min \left\{ \mathcal{O}(A_{q1}(\chi_{s1}), 1), \dots, \mathcal{O}(A_{qn}(\chi_{sn}), 1) \right\} \right) \text{ pela Proposição 1} \\
&\leq \min \left\{ M \left( \mathcal{O}(A_{q1}(\chi(11)), 1), \dots, \mathcal{O}(A_{q1}(\chi(s1)), 1) \right), \dots, \right. \\
&\quad \left. M \left( \mathcal{O}(A_{qn}(\chi(1n)), 1), \dots, \mathcal{O}(A_{qn}(\chi(sn)), 1) \right) \right\} \text{ pelo Corolário 2} \\
&= \min \left\{ \mathcal{O} \left( \left( \overset{s}{A_{qi}} \right)_1, U \right), \dots, \mathcal{O} \left( \left( \overset{s}{A_{qi}} \right)_n, U \right) \right\} \text{ pelas Equações (1) e (20)}
\end{aligned}$$

### 3.3 Mecanismo de Raciocínio de um Sistema de Classificação Baseado em Regras *Fuzzy*

Nos SCBRFs, a escolha apropriada da função de agregação a ser usada desempenha um papel crucial. Esta função de agregação determina o comportamento do MRF, uma vez que a informação local dada por cada regra *fuzzy* é agregada para fornecer informação global, que está associada a cada classe do problema. Finalmente, o exemplo é atribuído à classe que possui a informação global máxima, ou seja, a abordagem vencedora que seleciona o rótulo da classe associado à regra que fornece o maior grau de ativação da regra.

Considere um conjunto de  $\mathbb{M}$  classes  $\{C_1, \dots, C_{\mathbb{M}}\}$  e um conjunto de  $p$  regras  $\{R_1, \dots, R_p\}$ , com  $\mathbb{M} \leq p$ , dado como:

Regra  $R_q$ : Se  $\chi_1$  é  $A_{q1}$  e  $\dots$  e  $\chi_n$  é  $A_{qn}$  então Classe  $C_q$  com  $Cnf_q$ ,

onde  $q \in \{1, \dots, p\}$ ,  $C_q \in \{C_1, \dots, C_{\mathbb{M}}\}$  representa a classe da regra  $R_q$ , e  $Cnf_q$  é o grau de confiança da regra  $R_q$ , que é calculado pela aplicação do Algoritmo 1.

Então, dado o esquema:

Regra  $R_1$ : Se  $\chi_1$  é  $A_{11}$  e  $\dots$  e  $\chi_n$  é  $A_{1n}$  então Classe  $C_1$  com  $Cnf_1$

⋮

Regra  $R_p$ : Se  $\chi_1$  é  $A_{p1}$  e  $\dots$  e  $\chi_n$  é  $A_{pn}$  então Classe  $C_p$  com  $Cnf_p$

Fato:  $\chi_1$  é  $A'_1$  e  $\dots$  e  $\chi_n$  é  $A'_n$

apresentamos o Algoritmo 2 para calcular a classe  $C \in \{C_1, \dots, C_M\}$  para o qual o fato pertence, levando em consideração diferentes sobreposições entre os fatos e os antecedentes das regras. Representamos essas sobreposições por meio de diferentes índices de *overlap*. Esse método de raciocínio (usando diferentes índices de *overlap*) é o que distingue o Algoritmo 2 proposto dos clássicos (veja ISHIBUCHI et al (2005)), que demonstramos ser exemplos particulares do Algoritmo 2.

---

### Algoritmo 2

**Entrada:** Um conjunto de regras  $R_j$ , com  $j \in \{1, \dots, p\}$ , um conjunto de classes  $\{C_1, \dots, C_M\}$ , com  $M \leq p$  e um Fato  $\{(A'_1, A_{j1}), \dots, (A'_n, A_{jn})\}$ .

**Saída:** Uma classe  $C \in \{C_1, \dots, C_M\}$ .

- 1: Selecionar uma t-norma  $T$  e uma função de agregação  $M$ ;
  - 2: Selecionar um conjunto de  $S$  índices de *overlap*  $\{\mathcal{O}_1, \dots, \mathcal{O}_S\}$ ;
  - 3: Executar  $r$  vezes o Algoritmo 1. Cada regra  $R_j$  atribuiu um  $r$ -tupla de graus de confiança (e suporte):  $(Cnf_1^j, \dots, Cnf_r^j)$ ;
  - 4: **for**  $C = C_1$  to  $C_M$  **do**
  - 5:   Selecionar o conjunto de regras  $R_j$  que atribuiu a classe  $C$ ;
  - 6:    $\alpha =$  número de regras que atribuíram a classe  $C$ ;
  - 7:   **for**  $t = 1$  to  $r$  **do**
  - 8:     **for**  $L = 1$  to  $S$  **do**
  - 9:      **for**  $j = 1$  to  $\alpha$  **do**
  - 10:        Calcular  $k_{jCnf_{\mathcal{O}_L}}^C = T(\mathcal{O}_L(A'_1, A_{j1}), \dots, \mathcal{O}_L(A'_n, A_{jn})) \cdot Cnf_t^j$ ;
  - 11:      **end for**
  - 12:        Calcular  $K_{Cnf_{\mathcal{O}_L}}^C = \max_{j=1}^{\alpha} k_{jCnf_{\mathcal{O}_L}}^C$ ;
  - 13:      **end for**
  - 14:        Calcular  $\mathbb{K}^C = M_{L=1}^{S \times r} K_{Cnf_{\mathcal{O}_L}}^C$ ;
  - 15:      **end for**
  - 16:   **end for**
  - 17: Seja  $C = \arg \max \mathbb{K}^C$ .
- 

É imediato que:

**Proposição 2** *Se o Algoritmo 2 considera os valores de  $Cnf_q$  dados pela Equação (16) ou Equação (17) e o índice de overlap  $\mathcal{O} : CF(U) \times CF(U) \rightarrow [0, 1]$  definido, para todo  $A, B \in CF(U)$ , tem-se*

$$\mathcal{O}(A, B) = \frac{1}{n} \sum_{i=1}^n A(u_i)B(u_i),$$

*então recuperamos o algoritmo de Ishibuchi (ISHIBUCHI et al, 2005).*

### 3.4 Usando Funções de Penalidade para Escolher a Melhor Classe em Conjuntos de Regras *Fuzzy*

Um fator chave no Algoritmo 2 é a função de agregação  $M$  selecionada para o Passo 1. Se, para um determinado sistema de classificação, executamos esse algoritmo várias vezes, cada um deles com uma função de agregação diferente, podemos ter resultados diferentes, gerando conjuntos de regras *fuzzy*.

Esse fato nos obriga a fornecer um sistema de consenso para selecionar a classe final resultante. Para esse sistema, adaptamos o método de tomada de decisão apresentado em (BUSTINCE et al, 2014), que usa para as funções de penalidade na fase de exploração  $P_{\nabla}$  definidas sobre um produto cartesiano de lattices.

Denotamos por  $C^*$  uma cadeia de elementos que pertencem  $[0, 1]$  e consideramos um produto Cartesiano  $\mathcal{L}_m^* = \underbrace{C^* \times \dots \times C^*}_m$ . Denotamos por  $B_{y_q}$  o conjunto *fuzzy* em  $U$  de modo que todos os valores de associação sejam iguais para  $y_q \in [0, 1]$ , que é,  $B_{y_q}(u) = y_q \in [0, 1]$ , para todo  $u \in U$ . Considere  $\vec{Y} = (y_1, \dots, y_m)$ ,  $\vec{B}_Y = (B_{y_1}, \dots, B_{y_m}) \in CF(U)^m$ .

**Teorema 8** *Seja  $K_i : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , com  $i = 1, \dots, n$ , funções semi-contínuas pouca quasi-convexa com um mínimo exclusivo em  $K_i(0) = 0$ , e  $\mathcal{D} : CF(U) \times CF(U) \rightarrow \mathbb{R}^+$  seja a distância entre conjuntos fuzzy, definidos, para todos  $X, Y \in CF(U)$ , por*

$$\mathcal{D}(A, B) = \sum_{i=1}^n |A(u_i) - B(u_i)|, \quad (21)$$

onde  $n = \text{card}(U)$ . Então o mapeamento  $P_{\nabla} : CF(U)^m \times \mathcal{L}_m^* \rightarrow \mathbb{R}^+$ , dado, para todo  $\vec{A} \in CF(U)^m, Y \in \mathcal{L}_m^*$ , por :

$$P_{\nabla}(\vec{A}, Y) = \sum_{q=1}^m K_q(\mathcal{D}(A_q, B_{y_q})) = \sum_{q=1}^m \sum_{q=1}^m K_q \left( \sum_{p=1}^n |A_q(u_p) - y_q| \right) \quad (22)$$

é uma função de penalidade definida sobre o produto cartesiano de lattices  $\mathcal{L}_m^{*(n+1)}$ .

Observe que **(P1)** e **(P2)** da Definição 12 segue da BUSTINCE et al (2014, Teorema 6). A prova de **(P3)** Definição 12 é devido à soma das funções semi-convexas e semi-contínuas inferiores  $K_i$ .

No Algoritmo 2, nós deixamos o usuário escolher  $S$  graus de confiança e  $r$  índices de *overlap*. Então, para cada classe  $C_i \in \mathcal{C} = \{C_1, \dots, C_M\}$  temos um conjunto de  $S \times r$  valores numéricos. Consideramos um produto cartesiano de vários lattices como classes em nosso problema. Além disso, cada lattice possui  $S \times r$  elementos:

$$C_1 \times \dots \times C_M = (K_{C_{nf1O_1}}^{C_1}, \dots, K_{C_{nfrO_S}}^{C_1}) \times \dots \times (K_{C_{nf1O_1}}^{C_M}, \dots, K_{C_{nfrO_S}}^{C_M}).$$

Considere um conjunto de  $\mathbb{M}$  classes e outro conjunto de  $\mathbb{M}$  funções de agregação. Em (BUSTINCE et al, 2014) é proposto um método de otimização, de modo que, a partir de todas as regras de agregação de  $\mathbb{M}$ , obtidas ao calcular as permutações com repetições das funções de agregação  $\mathbb{M}$ , podemos selecionar a tupla de tal forma que, quando aplicamos os componentes da tupla (na sua ordem) aos componentes  $S \times r$  da classe correspondente, obtemos a tupla  $\mathbb{M}$  - que minimiza a dissimilaridade entre as classes de entrada  $\mathbb{M}$  e a tupla de números considerada.

Modificamos (BUSTINCE et al, 2014, Algoritmo 1) para o nosso problema (Algoritmo 3), aplicando a função de penalidade adotada  $P_{\nabla}$  Equação (22) para calcular o desvio entre as saídas fornecidas pelas diferentes tuplas de funções de agregação e, em seguida, selecionar aquela que minimiza essa dissimilaridade.

---

### Algoritmo 3

---

**Entrada:** Usando o Algoritmo 2, calculamos:

$$(K_{C_{nf1}O_1}^{C_1}, \dots, K_{C_{nfr}O_S}^{C_1})$$

⋮

$$(K_{C_{nf1}O_1}^{C_M}, \dots, K_{C_{nfr}O_S}^{C_M})$$

**Saída:** Uma classe  $C \in \{C_1, \dots, C_M\}$ .

- 1: Selecionar uma função *penalty*  $P_{\nabla}$  definido sobre o produto cartesiano de  $\mathbb{M}$  lattices (Teorema 8);
- 2: Selecionar uma  $\mathbb{M}$ -tupla  $(M_1, \dots, M_{\mathbb{M}})$  de funções de agregação idempotentes;
- 3: Calcular todas as permutações com repetição  $(M_{\sigma((1))}, \dots, M_{\sigma(\mathbb{M})})$  da tupla  $(M_1, \dots, M_{\mathbb{M}})$ .

4: **for** cada permutação de passo 3 **do**

$$m_{\sigma(1)}^{C_1} = M_{\sigma(1)}(K_{C_{nf1}O_1}^{C_1}, \dots, K_{C_{nfr}O_S}^{C_1})$$

5: ⋮

$$m_{\sigma(\mathbb{M})}^{C_M} = M_{\sigma(\mathbb{M})}(K_{C_{nf1}O_1}^{C_M}, \dots, K_{C_{nfr}O_S}^{C_M})$$

6: Calcular a função de penalidade por

$$P_{\nabla}(C_1, \dots, C_M, (m_{\sigma(1)}^{C_1}, \dots, m_{\sigma(\mathbb{M})}^{C_M})); \quad (23)$$

7: **end for**

8: Pegue a tupla de agregações  $(M_{m(\sigma(1))}^{C_1}, \dots, M_{m(\sigma(\mathbb{M}))}^{C_M})$ , que, com os valores obtidos no passo 5, minimiza a Eq. (23);

9: Calcular

$$\mathbb{K}^{C_1} = M_{m(\sigma(1))}^{C_1}(K_{C_{nf1}O_1}^{C_1}, \dots, K_{C_{nfr}O_S}^{C_1})$$

⋮

$$\mathbb{K}^{C_M} = M_{m(\sigma(\mathbb{M}))}^{C_M}(K_{C_{nf1}O_1}^{C_M}, \dots, K_{C_{nfr}O_S}^{C_M})$$

10: Seja  $C = \arg \max_{\{C_1, \dots, C_M\}} \mathbb{K}^C$

---

É imediato que:

**Proposição 3** *Seja  $\mathcal{M} = (M_1, \dots, M_{\mathbb{M}})$  uma tupla de funções de agregação idempotente. Se rodarmos o Algoritmo 2 várias vezes com diferentes funções de agregação idempotentes  $M \in \mathcal{M}$ , quando rodarmos o algoritmo de consenso (Algoritmo 3), tem-se que:*

- (i) *Se o  $M_1 = \dots = M_{\mathbb{M}}$ , então o resultado é dado pelo Algoritmo 2.*
- (ii) *Se, para todo  $i = 1, \dots, \mathbb{M}$ ,  $M_i$  é a média aritmética, então o Algoritmo 3 é o algoritmo de votação ponderada na tomada de decisão ((BUSTINCE et al, 2012)).*

**Observação 2** *Note-se que um raciocínio análogo sobre o funcionamento do Algoritmo 2 com diferentes funções de agregação  $M$  e a execução do Algoritmo 3 também podem ser feitas usando diferentes  $t$ -normas na Etapa 1 do Algoritmo 2. Não desenvolvemos este caso, pois é semelhante ao de diferentes funções de agregação  $M$ .*

### 3.5 Experimento

O exemplo consiste em:

1. Três classes, denominadas,  $C_1, C_2$  e  $C_3$ .
2. Duas variáveis linguísticas (atributos), denominadas,  $\chi_1$  (índice de massa corporal - IMC) e  $\chi_2$  (idade). Os valores  $\chi_1$  e  $\chi_2$  são qualificados pelos termos lingüísticos baixo, médio e alto, como representado na Figura 1. Considerando o gráfico no topo da figura, para a variável linguística  $\chi_1$ , os valores reais dos índices de massa corporal são marcados no eixo  $X$ , variando de 14 a 42. Analogamente, no gráfico na parte inferior da figura, para a variável linguística  $\chi_2$ , os valores reais das marcações estão no eixo  $X$  e variam de 30 a 70. Os graus de associação são marcados no eixo  $Y$ , variando de 0 a 1, o que refletirá os valores dos termos lingüísticos de tais variáveis nas funções de associação. Por exemplo, um índice de massa corporal de 35 é fuzzificado como: “ $\chi_1$  baixo com pertinência 0, médio com pertinência 0.5 e alto com pertinência 0.5”.
3. Vinte e cinco exemplos ou padrões (isto é,  $m = 25$ ), mostrados na Tabela 5. A figura 2 mostra os exemplos usados para aprender as regras *fuzzy* e o novo exemplo a ser classificado assim que as regras forem aprendidas. Para o processo de aprendizagem pode-se observar que temos quatro exemplos de classes  $C_1$  (sinal de adição), 10 exemplos da classe  $C_2$  (círculos) e 11 exemplos da classe  $C_3$  (triângulos).
4. Cinco regras aprendidas (R1-R5), como mostra a figura 2), pelos dados da Tabela 5 e usado alguns dos métodos em (CHI et al, 1996; SANZ et al, 2011, 2012):

Regra  $R_1$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Baixo então Classe  $C_1$  com  $Cnf_1 = ?$

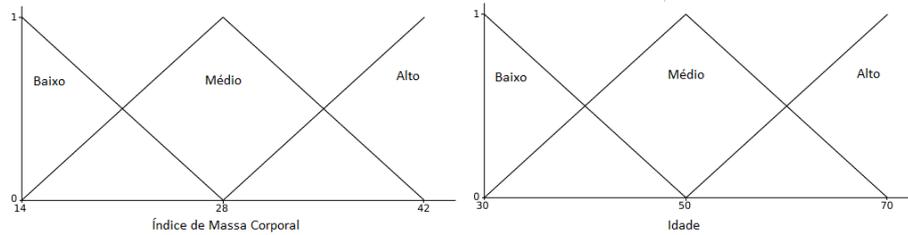


Figura 1: Termos Linguísticos para IMC e Idade

Tabela 5: Exemplos usados para treinamento

Id Exemplo	$\chi_1 = \text{IMC}$	$\chi_2 = \text{Idade}$	Classe
1	28	32	1
2	34	32	1
3	33	35	1
4	29	37	1
5	34	38	2
6	30	47	2
7	33	48	2
8	31	52	2
9	29	55	3
10	34	55	3
11	30	63	2
12	33	62	3
13	29	64	2
14	32	62	3
15	30	68	2
16	36	48	2
17	39	49	3
18	36	50	2
19	37	55	3
20	41	57	3
21	38	63	2
22	37	65	3
23	41	64	3
24	39	68	3
25	36	69	3

Regra  $R_2$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Médio então Classe  $C_2$  com  $Cnf_2 = ?$

Regra  $R_3$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Alto então Classe  $C_2$  com  $Cnf_3 = ?$

Regra  $R_4$ : Se  $\chi_1$  é Alto e  $\chi_2$  é Médio então Classe  $C_3$  com  $Cnf_4 = ?$

Regra  $R_5$ : Se  $\chi_1$  é Alto e  $\chi_2$  é Alto então Classe  $C_3$  com  $Cnf_5 = ?$

Da análise da Tabela 5, foram deduzidos 25 exemplos que estão na forma:

$$\chi_l = (\chi_{l1}, \chi_{l2}),$$

com  $l = 1, \dots, 25$ , onde  $\chi_{l1}$  representa o valor da variável  $\chi_1$  (IMC) e  $\chi_{l2}$  (Idade) é o valor da variável  $\chi_2$  no exemplo  $l$ .

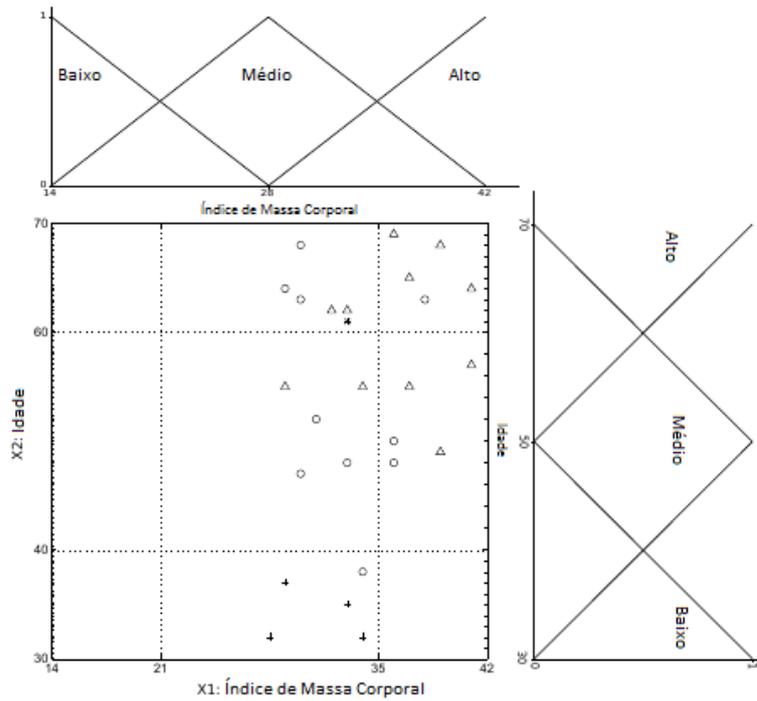


Figura 2: Regras

### 3.5.1 Avaliando os Graus de Confiança para Cada Regra

Construindo os graus de confiança  $Cnf_q$  para cada regra ( $q = 1, \dots, 5$ ). Os graus de pertinência dos valores da Tabela 5 para os conjuntos representados na Figura 1 são dados na Tabela 6.

Tendo em conta a Tabela 5, a distribuição das regras entre as três classes e Tabela 6, calculamos os graus de confiança das regras pelo Algoritmo 1 como segue:

**Passo 1:** Primeiro considere a t-norma do produto  $T = T_P$  e o índice de *overlap*  $\mathcal{O}_\pi$  (dado na Tabela 2). Temos  $m = 25$  exemplos, então  $A \in CF(U)$ , temos que:

$$\mathcal{O}_\pi(A, U) = \frac{1}{25} \sum_{u \in U} A(u) \cdot 1.$$

**Passo 2:** Classifique os exemplos nas três classes. Por exemplo, os quatro primeiros exemplo são  $C_1$  (veja Tabela 5).

**Passo 3:** Para  $q = 1$  até  $p = 5$  (isto é, para todas as regras), e começando com  $q = 1$ :

**Passo 4:** Levando  $s = 4$ , relacionado com os primeiros quatro exemplos pertencentes à classe  $C_1$  (veja Tabela 5);

**Passo 5:** Para  $j = 1$  até  $s = 4$ :

Tabela 6: Graus de pertinência *fuzzy*

Id Exemplo	$\chi_1 = \text{IMC}$			$\chi_2 = \text{Idade}$			Classe
	Baixo	Médio	Alto	Baixo	Médio	Alto	
1	0.000	1.000	0.000	0.900	0.100	0.000	1.000
2	0.000	0.571	0.429	0.900	0.100	0.000	1.000
3	0.000	0.643	0.357	0.750	0.250	0.000	1.000
4	0.000	0.929	0.071	0.650	0.350	0.000	1.000
5	0.000	0.571	0.429	0.600	0.400	0.000	2.000
6	0.000	0.857	0.143	0.150	0.850	0.000	2.000
7	0.000	0.643	0.357	0.100	0.900	0.000	2.000
8	0.000	0.786	0.214	0.000	0.900	0.100	2.000
9	0.000	0.929	0.071	0.000	0.750	0.250	3.000
10	0.000	0.571	0.429	0.000	0.750	0.250	3.000
11	0.000	0.857	0.143	0.000	0.350	0.650	2.000
12	0.000	0.643	0.357	0.000	0.400	0.600	3.000
13	0.000	0.929	0.071	0.000	0.300	0.700	2.000
14	0.000	0.741	0.286	0.000	0.400	0.600	3.000
15	0.000	0.857	0.143	0.000	0.100	0.900	2.000
16	0.000	0.429	0.571	0.100	0.900	0.000	2.000
17	0.000	0.241	0.786	0.050	0.950	0.000	3.000
18	0.000	0.429	0.571	0.000	1.000	0.000	2.000
19	0.000	0.357	0.643	0.000	0.750	0.250	3.000
20	0.000	0.071	0.929	0.000	0.650	0.350	3.000
21	0.000	0.286	0.714	0.000	0.350	0.650	2.000
22	0.000	0.357	0.643	0.000	0.250	0.750	3.000
23	0.000	0.071	0.929	0.000	0.300	0.700	3.000
24	0.000	0.214	0.786	0.000	0.100	0.900	3.000
25	0.000	0.429	0.571	0.000	0.050	0.950	3.000

**Passo 6:** Calcular o grau de associação, levando em conta os valores da Tabela 6, começando por  $j = 1$ :

$$c_1(\chi_1) = T_P(A_{11}(\chi_{11}), A_{12}(\chi_{12})) = A_{11}(28) \cdot A_{12}(32) = 1 \cdot 0.900 = 0.900$$

$$c_2(\chi_2) = T_P(A_{11}(\chi_{21}), A_{12}(\chi_{22})) = A_{11}(34) \cdot A_{12}(32) = 0.571 \cdot 0.900 = 0.5143$$

$$c_3(\chi_3) = T_P(A_{11}(\chi_{31}), A_{12}(\chi_{32})) = A_{11}(33) \cdot A_{12}(35) = 0.643 \cdot 0.750 = 0.4821$$

$$c_4(\chi_4) = T_P(A_{11}(\chi_{41}), A_{12}(\chi_{42})) = A_{11}(29) \cdot A_{12}(37) = 0.929 \cdot 0.650 = 0.6036$$

onde  $A_{11}$  e  $A_{12}$  correspondem aos conjuntos *fuzzy* que definem os termos lingüísticos Médio e Baixo, respectivamente, na regra  $R_1$ .

**Passo 7:** Construimos o conjunto *fuzzy* em  $U$ :

$$\begin{aligned} \mathbb{C}_{14} &= \{(u_1, c_1(\chi_1)), (u_2, c_2(\chi_2)), (u_3, c_3(\chi_3)), (u_4, c_4(\chi_4))\} \\ &= \{(u_1, 0.9000), (u_2, 0.5143), (u_3, 0.4821), (u_4, 0.6036), (u_5, 0), \dots, (u_{25}, 0)\} \end{aligned}$$

**Passo 8:** Para  $l = 1$  até  $m = 25$ , começando com  $l = 1$ :

**Passo 9:** Calculamos:

$$\begin{aligned}
c_5(\chi_5) &= T_p(A_{11}(\chi_{51}), A_{12}(\chi_{52})) = A_{11}(34) \cdot A_{12}(38) = 0.571 \cdot 0.600 = 0.3429 \\
c_6(\chi_6) &= T_p(A_{11}(\chi_{61}), A_{12}(\chi_{62})) = A_{11}(30) \cdot A_{12}(47) = 0.857 \cdot 0.150 = 0.1286 \\
c_7(\chi_7) &= T_p(A_{11}(\chi_{71}), A_{12}(\chi_{72})) = A_{11}(33) \cdot A_{12}(48) = 0.643 \cdot 0.100 = 0.0643 \\
c_8(\chi_8) &= T_p(A_{11}(\chi_{81}), A_{12}(\chi_{82})) = A_{11}(31) \cdot A_{12}(52) = 0.786 \cdot 0.000 = 0.0000 \\
c_9(\chi_9) &= T_p(A_{11}(\chi_{91}), A_{12}(\chi_{92})) = A_{11}(29) \cdot A_{12}(55) = 0.929 \cdot 0.000 = 0.0000 \\
c_{10}(\chi_{10}) &= T_p(A_{11}(\chi_{101}), A_{12}(\chi_{102})) = A_{11}(34) \cdot A_{12}(55) = 0.571 \cdot 0.000 = 0.0000 \\
c_{11}(\chi_{11}) &= T_p(A_{11}(\chi_{111}), A_{12}(\chi_{112})) = A_{11}(30) \cdot A_{12}(63) = 0.857 \cdot 0.000 = 0.0000 \\
c_{12}(\chi_{12}) &= T_p(A_{11}(\chi_{121}), A_{12}(\chi_{122})) = A_{11}(33) \cdot A_{12}(62) = 0.643 \cdot 0.000 = 0.0000 \\
c_{13}(\chi_{13}) &= T_p(A_{11}(\chi_{131}), A_{12}(\chi_{132})) = A_{11}(29) \cdot A_{12}(64) = 0.929 \cdot 0.000 = 0.0000 \\
c_{14}(\chi_{14}) &= T_p(A_{11}(\chi_{141}), A_{12}(\chi_{142})) = A_{11}(32) \cdot A_{12}(62) = 0.714 \cdot 0.000 = 0.0000 \\
c_{15}(\chi_{15}) &= T_p(A_{11}(\chi_{151}), A_{12}(\chi_{152})) = A_{11}(30) \cdot A_{12}(68) = 0.857 \cdot 0.000 = 0.0000 \\
c_{16}(\chi_{16}) &= T_p(A_{11}(\chi_{161}), A_{12}(\chi_{162})) = A_{11}(36) \cdot A_{12}(48) = 0.429 \cdot 0.100 = 0.0429 \\
c_{17}(\chi_{17}) &= T_p(A_{11}(\chi_{171}), A_{12}(\chi_{172})) = A_{11}(39) \cdot A_{12}(49) = 0.214 \cdot 0.050 = 0.0107 \\
c_{18}(\chi_{18}) &= T_p(A_{11}(\chi_{181}), A_{12}(\chi_{182})) = A_{11}(36) \cdot A_{12}(50) = 0.429 \cdot 0.000 = 0.0000 \\
c_{19}(\chi_{19}) &= T_p(A_{11}(\chi_{191}), A_{12}(\chi_{192})) = A_{11}(37) \cdot A_{12}(55) = 0.357 \cdot 0.000 = 0.0000 \\
c_{20}(\chi_{20}) &= T_p(A_{11}(\chi_{201}), A_{12}(\chi_{202})) = A_{11}(41) \cdot A_{12}(57) = 0.071 \cdot 0.000 = 0.0000 \\
c_{21}(\chi_{21}) &= T_p(A_{11}(\chi_{211}), A_{12}(\chi_{212})) = A_{11}(38) \cdot A_{12}(63) = 0.286 \cdot 0.000 = 0.0000 \\
c_{22}(\chi_{22}) &= T_p(A_{11}(\chi_{221}), A_{12}(\chi_{222})) = A_{11}(37) \cdot A_{12}(65) = 0.357 \cdot 0.000 = 0.0000 \\
c_{23}(\chi_{23}) &= T_p(A_{11}(\chi_{231}), A_{12}(\chi_{232})) = A_{11}(41) \cdot A_{12}(64) = 0.071 \cdot 0.000 = 0.0000 \\
c_{24}(\chi_{24}) &= T_p(A_{11}(\chi_{241}), A_{12}(\chi_{242})) = A_{11}(39) \cdot A_{12}(68) = 0.214 \cdot 0.0000 = 0.0000 \\
c_{25}(\chi_{25}) &= T_p(A_{11}(\chi_{251}), A_{12}(\chi_{252})) = A_{11}(36) \cdot A_{12}(69) = 0.429 \cdot 0.000 = 0.0000
\end{aligned}$$

**Passo 10:** Construimos o conjunto *fuzzy*  $U$ :

$$\begin{aligned}
\mathbb{C}_{1,25} &= \{(u_1, c_1(\chi_1)), \dots, (u_{25}, c_m(\chi_{25}))\} \\
&= \{(u_1, 0.9000), (u_2, 0.5143), (u_3, 0.4821), (u_4, 0.6036), (u_5, 0.3429), \\
&\quad (u_6, 0.1286), (u_7, 0.0643), (u_8, 0.0000), (u_9, 0.0000), (u_{10}, 0.0000), \\
&\quad (u_{11}, 0.0000), (u_{12}, 0.0000), (u_{13}, 0.0000), (u_{14}, 0.0000), (u_{15}, 0.0000), \\
&\quad (u_{16}, 0.0429), (u_{17}, 0.0107), (u_{18}, 0.0000), (u_{19}, 0.0000), (u_{20}, 0.0000) \\
&\quad (u_{21}, 0.0000), (u_{22}, 0.0000), (u_{23}, 0.0000), (u_{24}, 0.0000), (u_{25}, 0.0000)\}.
\end{aligned}$$

**Passo 11:** Usando a Eq. (14), calculamos:

Tabela 7:  $Cnf$  construído a partir de diferentes índices de *overlap*

Regra	$O_\pi$	$O_\surd$	$O_Z$	$O_{rat}$
1	0.8092	0.6753	1	0.6867
2	0.5617	0.4964	1	0.4993
3	0.5024	0.4160	1	0.4290
4	0.5579	0.5164	1	0.5188
5	0.8169	0.7586	1	0.7598

$$\begin{aligned}
& Cnf(R_1) \\
&= \frac{\mathcal{O}(C_{14}, U)}{\mathcal{O}(C_{125}, U)} \\
&= \frac{\frac{1}{25} \sum_{i=1}^4 c_i(\chi_i)}{\frac{1}{25} \sum_{i=1}^{25} c_i(\chi_i)} \\
&= \frac{\frac{1}{25} (0.90000 + 0.5143 + 0.4821 + 0.6036)}{\frac{1}{25} (0.9000 + 0.5143 + 0.4821 + 0.6036 + 0.3429 + 0.1286 + 0.0643 + 0.0429 + 0.0107)} \\
&= \frac{2.5}{3.0893} \\
&= 0.8092
\end{aligned}$$

**Passos seguintes:** Repetindo as etapas anteriores para  $q = 2, 3, 4, 5$ , temos os seguintes resultados para outras regras:  $Cnf(R_2) = 0.5671$ ,  $Cnf(R_3) = 0.5024$ ,  $Cnf(R_4) = 0.5579$ ,  $Cnf(R_5) = 0.8169$  (veja a segunda coluna da Tabela 7).

Se no passo 1 do Algoritmo 1, tomamos os índices de *overlap* da Tabela 3, então obtemos os graus de confiança  $Cnf$  presentes nas colunas 3, 4 e 5 respectivamente na Tabela 7.

### 3.5.2 Inferência

Dado o conjunto de regras, com seus respectivos graus de confiança:

Regra  $R_1$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Baixo então Classe  $C_1$  com  $Cnf_1$

Regra  $R_2$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Médio então Classe  $C_2$  com  $Cnf_2$

Regra  $R_3$ : Se  $\chi_1$  é Médio e  $\chi_2$  é Alto então Classe  $C_2$  com  $Cnf_3$

Regra  $R_4$ : Se  $\chi_1$  é Alto e  $\chi_2$  é Médio então Classe  $C_3$  com  $Cnf_4$

Regra  $R_5$ : Se  $\chi_1$  é Alto e  $\chi_2$  é Alto então Classe  $C_3$  com  $Cnf_5$

e o seguinte fato:

Índice de massa corporal = 33 e Idade = 61

Pretendemos determinar a qual classe este fato pertence, usando Algoritmo 2. Na Figura 2 o novo exemplo a ser classificado (IMC=33 e Idade=61) é representado com o asterisco e de acordo com sua posição pode-se inferir que pertence à classe três, já que está rodeado de exemplos dessa classe.

Agora, da Figura 1:

1. Construímos os conjuntos  $A'_1$  e  $A'_2$  como segue:

$$A'_1 = \{(14, 0), (15, 0), \dots, (32, 0), (33, 1), (34, 0), \dots, (42, 0)\}$$

$$A'_2 = \{(30, 0), (31, 0), \dots, (60, 0), (61, 1), (62, 0), \dots, (70, 0)\}$$

Por uma questão de simplicidade, apresentamos esta construção. Na verdade, podemos escolher diferentes construções, dependendo da aplicação em que estamos trabalhando.

2. Calcular os graus de pertinência da variável linguística IMC (33) para o conjunto *fuzzy* Baixo, Médio e Alto, que representa os termos linguísticos que qualificam a variável linguística IMC. Os resultados são: Baixo(33)= 0.000, Médio(33)= 0.6429 e Alto(33)= 0.3571.
3. Analogamente, calculamos os graus de pertinência da variável linguística Idade (61), para os conjuntos *fuzzy* Baixo, Médio e Alto, obtemos os resultados: Baixo(61)= 0.000; Médio(61)= 0.450 e Alto(61)= 0.550.

Executando o Algoritmo 2 para aprendizagem, temos:

**Passo 1:** Consideramos a t-norma  $T = T_P$  e a função de agregação  $M = \max$

**Passo 2:** Consideramos o conjunto de  $S = 3$  índices de *overlap*, a saber,  $\{\mathcal{O}_\pi, \mathcal{O}_\surd, \mathcal{O}_{rat}\}$  (Tabela 3)

**Passo 3:** Os graus de confiança  $Cnf_j$  ( $j = 1, \dots, 5$ ) são calculados com o Algoritmo 1 e representados nas colunas 2, 3 e 5 da Tabela 7

**Passo 4:** Aplicado a t-norma nos índice de *overlap* de cada regra e após multiplicado pelo grau de confiança da regra, obtendo o resultado da Tabela 8

**Passo 5:** Aplicar a agregação do máximo.

Uma vez que o Algoritmo 2 foi executado com a função de agregação  $M_1 = \max = m_1$ , executamos este algoritmo duas vezes mais: uma com a agregação  $M_2$ , dado pelo operador OWA correspondente ao quantificador *pelo menos metade*, isto é, considerando o vetor de pesos  $\omega = (0.33, 0.33, 0.33, 0, 0, 0)$ , e outro com  $M_3$  dado pelo operador OWA correspondente ao quantificador *o maior número possível*, ou seja, considerando o vetor de peso  $\omega = (0, 0, 0, 0.33, 0, 33, 0, 33)$ .

Em seguida, executamos o algoritmo de consenso, Algoritmo 3, com a função de penalidade  $P_\nabla$  do Teorema 8, definida, nesse contexto, por

$$P_\nabla(C_1, \dots, C_M, (m_{\sigma(1)}^{C_1}, \dots, m_{\sigma(M)}^{C_M})) = \sum_{i=1}^M \left( \sum_{j=1, \dots, r; L=1, \dots, S} |K_{Cnf_j \mathcal{O}_L}^{C_i} - m_{\sigma(i)}^{C_i}| \right)^2 \quad (24)$$

e as tuplas construídas calculando as permutações com a repetição de  $(M_1, M_2, M_3)$ . Então temos 27 combinações possíveis das 3 funções de agregação.

Na Tabela 8 apresentamos os valores numéricos obtidos com a função penalidade utilizada para cada uma das 3-tuplas consideradas. Observe que o valor mínimo é de 0,3366, correspondendo à combinação de agregações:  $(M_1, M_3, M_3)$ . Observe que o mesmo valor também é obtido com as combinações  $(M_2, M_3, M_3)$  e  $(M_3, M_3, M_3)$ , já que a regra da classe  $C_1$  não foi ativada.

Portanto, temos que  $\mathbb{K}^{C_1} = 0$ ,  $\mathbb{K}^{C_2} = 0.1028$ ,  $\mathbb{K}^{C_3} = 0.2318$ . Pelo Passo 10 do Algoritmo 3 a classe resultante é  $C_3$ .

### 3.6 Análise do Experimento

A Tabela 8 mostra o resultado obtido com o Algoritmo 2 para o experimento, analisando por colunas, vemos que cada coluna central corresponde aos dados obtidos com o algoritmo clássico de classificação, fixando um grau de confiança para as regras e escolhendo o conjunto de índices de *overlap* em Passo 2 como o conjunto que consiste em um único *overlap*. Se analisarmos a Tabela 8 por linhas, o resultado é aquele obtido por meio do Algoritmo 2 com dois graus de confiança e três índices de *overlap*.

Então a Tabela 8 e o Algoritmo 2 nos permite estudar os resultados da classificação de um dado fato de duas maneiras diferentes: por meio do algoritmo clássico (por colunas) ou por meio de uma família de diferentes medidas para representar a sobreposição entre os antecedentes das regras e o fato considerado (por linhas).

Do nosso ponto de vista, devemos usar Algoritmo 2 e entender Tabela 8 por linhas, pois na classificação sempre existem diferentes tipos de interações entre fatos e antecedentes, e devemos levar em conta todos deles. Isso não acontece nos algoritmos clássicos, que medem apenas a sobreposição de uma maneira.

Da análise da primeira coluna na Tabela 8 deduzimos que o fato que estamos considerando neste exemplo pertence a  $C_2$ . No entanto, a partir da análise das outras colunas (exceto a quarta), deduzimos que o fato pertence à classe  $C_3$ .

Como não houve um consenso na classificação, recorreremos ao uso da função *penalty* sobre o produto cartesiano de *lattices*. Como foram escolhidas 3 três funções de agregação, houve uma permutação de 27 itens para 3 classes, que gera uma ganho considerável no processamento, tendo em vista que o experimento foi realizado com uma base de dados pequena.

Tabela 8: Resultados do Algoritmo 2

Classe	$Cnf = Cnf$ Construido pelo Algoritmo 1 usando $\mathcal{O}_\pi$			$Cnf = Cnf$ Construido pelo Algoritmo 1 usando $\mathcal{O}_{\pi'}$			$M = \max$	$C = \text{arg max} \mathbb{K}^C$
	$\mathcal{O}_\pi$	$\mathcal{O}_\gamma$	$\mathcal{O}_{\pi'}$	$\mathcal{O}_\pi$	$\mathcal{O}_\gamma$	$\mathcal{O}_{\pi'}$		
C1	$K_{Cnf_{O_\pi}}^{C1} = 0$	$K_{Cnf_{O_\gamma}}^{C1} = 0$	$K_{Cnf_{O_{\pi'}}}^{C1} = 0$	$K_{Cnf_{O_\pi}}^{C1} = 0$	$K_{Cnf_{O_\gamma}}^{C1} = 0$	$K_{Cnf_{O_{\pi'}}}^{C1} = 0$	$\mathbb{K}^{C1} = 0$	
C2	$K_{Cnf_{O_\pi}}^{C2} = V(0.1625, 0.1776)$	$K_{Cnf_{O_\gamma}}^{C2} = V(0.3021, 0.2987)$	$K_{Cnf_{O_{\pi'}}}^{C2} = V(0.2420, 0.2407)$	$K_{Cnf_{O_\pi}}^{C2} = V(0.1444, 0.1517)$	$K_{Cnf_{O_\gamma}}^{C2} = V(0.2685, 0.2551)$	$K_{Cnf_{O_{\pi'}}}^{C2} = V(0.2151, 0.2056)$	$\mathbb{K}^{C2} = 0.3021$	
C3	$K_{Cnf_{O_\pi}}^{C3} = V(0.0897, 0.1605)$	$K_{Cnf_{O_\gamma}}^{C3} = V(0.2237, 0.3621)$	$K_{Cnf_{O_{\pi'}}}^{C3} = V(0.1803, 0.2904)$	$K_{Cnf_{O_\pi}}^{C3} = V(0.0834, 0.1492)$	$K_{Cnf_{O_\gamma}}^{C3} = V(0.2080, 0.3367)$	$K_{Cnf_{O_{\pi'}}}^{C3} = V(0.1677, 0.2701)$	$\mathbb{K}^{C3} = 0.3621$	$C = C_3$
Classe	C2	C3	C3	C2	C3	C3		

Tabela 9: Resultados obtidos através do Algoritmo 3. Consenso

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.5717	0.3461	0.4242	0.5290	0.4394	0.4356	<b>0.3346</b>	0.4783	0.4669	0.4783	0.4356	0.4394	0.4669	0.3461	0.5717	<b>0.3346</b>	0.5290	0.4242	<b>0.3346</b>	0.5290	0.4356	0.4669	0.4242	0.5717	0.4783	0.4394	0.3461

## 4 APLICAÇÃO DAS FUNÇÕES DE (PRÉ)-AGREGAÇÃO BASEADAS NA INTEGRAL DE CHOQUET

Neste capítulo é apresentado a metodologia da aplicação das funções de (pré)-agregação generalizadas da integral de Choquet no algoritmo desenvolvido. Executamos esta proposta no mesmo exemplo do capítulo anterior e demonstramos os resultados obtidos, sem a utilização da função de penalidade.

### 4.1 Metodologia

Utilizar o mesmo Algoritmo 1 apresentado no capítulo 3, com a geração dos índices de *overlap* de cada regra. Implementar as funções de pré-agregação baseada na integral de Choquet no Algoritmo 2 e aplicar no mesmo conjunto de dados já apresentado, espera-se que a classificação já tenha um consenso sem a necessidade da execução do Algoritmo 3, que se refere a função *penalty* definida sobre as permutações das classes e funções de agregação.

Com a aplicação deste experimento espera-se ter um melhor resultado na tomada de decisão, pois a integral de Choquet agrega os valores levando em consideração a importância das entradas individuais, ou seja, uma entrada específica pode não ser importante por si só, mas se torna muito importante na presença de algumas outras entradas.

Adotamos a medida *fuzzy* uniforme que é simétrica, ou seja, a medida de qualquer subconjunto  $A$  só depende da cardinalidade de  $A$ .

Selecionamos algumas funções de pré-agregação baseadas da Integral de Choquet que nos artigos estudados obtiveram bons resultados nos comparativos. Conforme segue, a função de pré-agregação do produto de Hamacher  $T_{HP}$  (LUCCA et al, 2016), a  $CC$ -integral do Mínimo  $TM$  (LUCCA et al, 2016), a  $C_F$ -integral média  $F_{NA}$  e a  $C_F$ -integral não média  $F_{NA2}$  (LUCCA et al, 2018), a função  $C_{F1F2}$ -integral com o par de funções  $GM - T_{\underline{L}}$  e com o par de funções  $GM - F_{BPC}$  (LUCCA et al, 2018). As fórmulas de cada função estão descritas na Tabela 4.

## 4.2 Alteração do Algoritmo 2

Ao invés de utilizar uma função t-norma, descrita no Passo 1 do Algoritmo 2, alteramos para o uso de uma função de (pré)-agregação baseada na integral de Choquet.

---

### Algoritmo 4

---

**Entrada:** Um conjunto de regras  $R_j$ , com  $j \in \{1, \dots, p\}$ , um conjunto de classes  $\{C_1, \dots, C_M\}$ , com  $M \leq p$  e um Fato  $\{(A'_1, A_{j1}), \dots, (A'_n, A_{jn})\}$ .

**Saída:** Uma classe  $C \in \{C_1, \dots, C_M\}$ .

- 1: Selecionar uma função de (pré)-agregação  $P$  e uma função de agregação  $M$ ;
  - 2: Selecionar um conjunto de  $S$  índices de *overlap*  $\{\mathcal{O}_1, \dots, \mathcal{O}_S\}$ ;
  - 3: Executar  $r$  vezes o Algoritmo 1. Cada regra  $R_j$  atribuiu um  $r$ -tupla de graus de confiança (e suporte):  $(Cnf_1^j, \dots, Cnf_r^j)$ ;
  - 4: **for**  $C = C_1$  to  $C_M$  **do**
  - 5:   Selecionar o conjunto de regras  $R_j$  que atribuiu a classe  $C$ ;
  - 6:    $\alpha =$  número de regras que atribuíram a classe  $C$ ;
  - 7:   **for**  $t = 1$  to  $r$  **do**
  - 8:     **for**  $L = 1$  to  $S$  **do**
  - 9:      **for**  $j = 1$  to  $\alpha$  **do**
  - 10:        Calcular  $k_{jCnf_{\mathcal{O}_L}}^C = P(\mathcal{O}_L(A'_1, A_{j1}), \dots, \mathcal{O}_L(A'_n, A_{jn})) \cdot Cnf_t^j$ ;
  - 11:        **end for**
  - 12:        Calcular  $K_{Cnf_{\mathcal{O}_L}}^C = \max_{j=1}^{\alpha} k_{jCnf_{\mathcal{O}_L}}^C$ ;
  - 13:        **end for**
  - 14:        Calcular  $\mathbb{K}^C = \prod_{L=1}^{S \times r} K_{Cnf_{\mathcal{O}_L}}^C$ ;
  - 15:        **end for**
  - 16:    **end for**
  - 17: Seja  $C = \arg \max \mathbb{K}^C$ .
- 

## 4.3 Experimento

Tendo a mesma base dados apresentada no capítulo anterior, executamos o Algoritmo 1 e obtemos os valores da Tabela 7. Utilizando o mesmo fato, Índice de massa corporal = 33 e Idade = 61, executamos o Algoritmo 4 conforme segue:

**Passo 1:** Consideramos uma função de pré-agregação  $P = T_{HP}$  e a função de agregação  $M = \max$

**Passo 2:** Consideramos o conjunto de  $S = 3$  índices de *overlap*, a saber,  $\{\mathcal{O}_\pi, \mathcal{O}_\surd, \mathcal{O}_{rat}\}$  (Tabela 3)

**Passo 3:** Os graus de confiança  $Cnf_j$  ( $j = 1, \dots, 5$ ) são calculados com o Algoritmo 1 e representados nas colunas 2, 3 e 5 da Tabela 7

Tabela 10: Resultados do Algoritmo 4 com Função Pré-Agregação  $T_{HP}$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.2862, 0.2841)$	$K_{Cnf}^{C_2} = \vee(0.2544, 0.2426)$	$\mathbb{K}^{C_2} = 0.2862$	
C3	$K_{Cnf}^{C_3} = \vee(0.2110, 0.3418)$	$K_{Cnf}^{C_3} = \vee(0.1962, 0.3179)$	$\mathbb{K}^{C_3} = 0.3418$	$C = C_3$
Classe	C3	C3		

Tabela 11: Resultados do Algoritmo 4 com Função Pré-Agregação  $TM$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.2420, 0.2407)$	$K_{Cnf}^{C_2} = \vee(0.2151, 0.2056)$	$\mathbb{K}^{C_2} = 0.2420$	
C3	$K_{Cnf}^{C_3} = \vee(0.1860, 0.2904)$	$K_{Cnf}^{C_3} = \vee(0.1729, 0.2701)$	$\mathbb{K}^{C_3} = 0.2904$	$C = C_3$
Classe	C3	C3		

**Passo 4:** Aplicado a função de pré-agregação nos *overlaps* de cada regra e após multiplicado pelo grau de confiança da regra, obtemos o resultado da Tabela 10.

**Passo 5:** Aplicar a agregação do máximo resultando na classe C3.

#### 4.4 Análise do Experimento

Nas Tabelas 10, 11, 12, 13, 14 e 15 estão apresentados os valores correspondentes aos três *overlaps* do fato pré-agregados e multiplicados por cada índice de *overlap* ( $\mathcal{O}_\pi$  e  $\mathcal{O}_{rat}$ ).

Analizando as tabelas por linhas, o resultado é aquele obtido por meio do Algoritmo 4 com dois índices de *overlap*. O resultado das tabelas por coluna está com apenas um índice de *overlap*.

Usando o Algoritmo 4 e entendendo as Tabelas 10, 11, 12, 13, 14 e 15 percebemos que, tanto por linhas, como por colunas, houve um consenso para a tomada de decisão na classificação. Isso se deve ao uso da função de pré-agregação com base na integral de Choquet, que considera a relação entre os elementos que estão sendo agregados, fornecendo a relevância de um grupo por medidas *fuzzy*. O experimento realizado mostrou que houve uma melhora na agregação dos elementos para a escolha da classe, como era esperado, segundo os resultados em (LUCCA et al, 2016; LUCCA et al, 2016; DIMURO et al, 2017; LUCCA et al, 2018)

Tabela 12: Resultados do Algoritmo 4 com Função Pré-Agregação  $F_{NA}$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.3021, 0.2987)$	$K_{Cnf}^{C_2} = \vee(0.2685, 0.2551)$	$\mathbb{K}^{C_2} = 0.3021$	
C3	$K_{Cnf}^{C_3} = \vee(0.2237, 0.3621)$	$K_{Cnf}^{C_3} = \vee(0.2080, 0.3367)$	$\mathbb{K}^{C_3} = 0.3621$	$C = C_3$
Classe	C3	C3		

Tabela 13: Resultados do Algoritmo 4 com Função Pré-Agregação  $F_{NA2}$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.5131, 0.4894)$	$K_{Cnf}^{C_2} = \vee(0.4561, 0.4179)$	$\mathbb{K}^{C_2} = 0.5131$	
C3	$K_{Cnf}^{C_3} = \vee(0.4356, 0.6697)$	$K_{Cnf}^{C_3} = \vee(0.4051, 0.6229)$	$\mathbb{K}^{C_3} = 0.6697$	$C = C_3$
Classe	C3	C3		

Tabela 14: Resultados do Algoritmo 4 com Função Pré-Agregação  $C_{F_1F_2} GM - T_{\mathbf{L}}$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.7013, 0.6751)$	$K_{Cnf}^{C_2} = \vee(0.6234, 0.5765)$	$\mathbb{K}^{C_2} = 0.7013$	
C3	$K_{Cnf}^{C_3} = \vee(0.5526, 0.8722)$	$K_{Cnf}^{C_3} = \vee(0.5139, 0.8111)$	$\mathbb{K}^{C_3} = 0.8722$	$C = C_3$
Classe	C3	C3		

Tabela 15: Resultados do Algoritmo 4 com Função Pré-Agregação  $C_{F_1F_2} GM - F_{BPC}$ 

Classe	$Cnf = Cnf \mathcal{O}_\pi$	$Cnf = Cnf \mathcal{O}_{rat}$	$M = \max$	$C = \arg \max \mathbb{K}^C$
C1	$K_{Cnf}^{C_1} = 0$	$K_{Cnf}^{C_1} = 0$	$\mathbb{K}^{C_1} = 0$	
C2	$K_{Cnf}^{C_2} = \vee(0.6022, 0.5796)$	$K_{Cnf}^{C_2} = \vee(0.5353, 0.4949)$	$\mathbb{K}^{C_2} = 0.6022$	
C3	$K_{Cnf}^{C_3} = \vee(0.4927, 0.7686)$	$K_{Cnf}^{C_3} = \vee(0.4582, 0.7148)$	$\mathbb{K}^{C_3} = 0.7686$	$C = C_3$
Classe	C3	C3		

## 5 CONCLUSÃO

Na classificação existem diferentes tipos de interações entre fatos e antecedentes, e devemos levar em conta todos eles. Isso não acontece nos algoritmos clássicos, que medem apenas a sobreposição de uma maneira.

O uso do máximo como função de agregação no MRF para obter a informação global é muito comum na literatura. Isto é conhecido como o MRF da regra vencedora. No entanto, se considerarmos, para cada classe, apenas as informações fornecidas por uma única regra *fuzzy* que tenha a maior compatibilidade com o exemplo, as informações disponíveis fornecidas pelas regras *fuzzy* restantes do sistema são ignorados. Por isso a proposta do uso das funções de pré-agregação baseadas na integral de Choquet deste trabalho, para que cada elemento calculado fosse considerado na agregação.

Com este trabalho proporcionamos a interação entre os fatos e os antecedentes com o uso de três *overlaps*, e com a aplicação de algumas funções de (pré)-agregação, tornando a seleção da classe com maior relevância.

Introduzimos um método de tomada de decisão em conjuntos de sistemas de classificação baseados em regras *fuzzy* que usam esses conceitos e funções de (pré)-agregação. Para isso, desenvolvemos 4 algoritmos relacionados a SCBRFs:

1. O Algoritmo 1, que implementa um método para construir medidas de confiança e suporte, com base em índices de *overlap*, que, por sua vez, podem ser construídos por funções de *overlap*;
2. O Algoritmo 2 implementa um novo MRF para SCBRF, usando diferentes índices de *overlap* e funções de agregação para obter a classe selecionada através de um treinamento, generalizando alguns métodos clássicos;
3. O Algoritmo 3, que implementa o método de consenso para a classificação, baseado em funções de penalidade, quando necessário;
4. O Algoritmo 4, baseado no MRF do Algoritmo 2, usa diferentes índices de *overlap* e funções de (pré)-agregação para obter a classe selecionada através de um treinamento.

Como trabalhos futuros pretendemos aplicar em outras bases de dados os índices de *overlap* e as funções de (pré)-agregação baseadas na integral de Choquet para verificar se a descoberta da classe é feita sem a necessidade da função *penalty*, como ocorreu neste experimento, além disso estudar o impacto da medida *fuzzy* nas funções de pré-agregação para a tomada de decisão.

## REFERÊNCIAS

ALSINA, C.; SCHWEIZER, B. **Associative Functions: Triangular Norms and Copulas**, World Scientific Publishing Company, Singapore, 2006.

BARRENECHEA, E.; BUSTINCE, H.; FERNANDEZ, J.; PATERNAIN, D.; SANZ, J.A. **Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems**, *Axioms* 2, P. 208-223, 2013.

BEDREGAL, B. C.; DIMURO, G. P.; BUSTINCE, H.; BARRENECHEA, E. **New Results on Overlap and Grouping Functions**, [S.l.], *Information Sciences*, v. 249, p. 148-170, 2013.

BELIAKOV, G.; BUSTINCE, H.; CALVO, T. **A Practical Guide to Averaging Functions**, v 329, *Studies in Fuzziness and Soft Computing*, Springer, 2016.

BELIAKOV, G.; CALVO, T.; JAMES, S. **Consensus measures constructed from aggregation functions and fuzzy implications**, [S.l.], *Knowledge-Based Systems*, v. 55, p. 1-8, 2014.

BELIAKOV, G.; PRADERA, A.; CALVO, T. **Aggregation Functions: A Guide for Practitioners**, Berlin, Springer, 2007.

BUSTINCE, H.; BELIAKOV, G.; DIMURO, G. P.; BEDREGAL, B.; MESIAR, R. **On the definition of penalty functions in data aggregation**, [S.l.], *Fuzzy Sets and Systems*, v. 323, p. 1-18, 2016.

BUSTINCE, H.; FERNANDEZ, J.; MESIAR, R.; MONTERO, J.; ORDUNA, R. **Overlap Functions**, [S.l.], *Nonlinear Analysis: Theory, Methods & Applications*, v. 72, n. 3-4, p. 1488-1499, 2010.

BUSTINCE, H.; FERNANDEZ, J.; MESIAR, R.; MONTERO, J.; ORDUNA, R. **Overlap Index, Overlap Functions and Migrativity**, [S.l.], *Proceedings of IFSA/EUSFLAT Conference*, p. 300-305, 2009.

BUSTINCE, H.; PAGOLA, M.; MESIAR, R.; HLLERMEIER, E.; HERRERA, F. **Grouping, Overlaps, and Generalized Bientropic Functions for Fuzzy Modeling**, [S.l.], IEEE Transactions on Fuzzy Systems, v. 20, n. 3, p. 405-415, 2012.

BUSTINCE, H.; BARRENECHEA, E.; CALVO, T.; JAMES, S.; BELIAKOV, G. **Consensus in multi-expert decision making problems using penalty functions defined over a cartesian product of lattices**, Information Fusion, V. 17, p. 56-64, 2014.

CALVO, T.; BELIAKOV, G. **Aggregation functions based on penalties**, [S.l.], Fuzzy Sets and Systems, v. 161, n. 10, p. 1420-1436, 2010.

CALVO, T.; MESIAR, R.; YAGER, R. R. **Quantitative weights and aggregation**, [S.l.], IEEE Transactions on Fuzzy Systems, v. 12, n. 1, p. 62-69, 2004.

CHI, Z.; YAN, H; PHAM, T. **Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition**, World Scientific, Singapore, 1996.

CHOQUET, G. **Theory of capacities**, [S.l.], l'Institut Fourier, v. 5, p. 131-295, 1953.

DENNEBERG, D. **Non Additive Measure and Integral**, Dordrecht, Kluwer Academic Publishers, 1994.

DIMURO, G. P.; BEDREGAL, B. **Archimedean overlap functions: The ordinal sum and the cancellation, idempotency and limiting properties**, [S.l.], Fuzzy Sets and Systems, v. 252, p. 39-54, 2014.

DIMURO, G. P.; BEDREGAL, B. **On residual implications derived from overlap functions**, [S.l.], Information Sciences, v.312, p. 78-88, 2015.

DIMURO, G. P.; BEDREGAL, B. **On the laws of contraposition for residual implications derived from overlap functions**, Los Alamitos, IEEE, p. 1-7, 2015.

DIMURO, G. P.; BEDREGAL, B.; BUSTINCE, H.; ASIAN, M. J.; MESIAR, R. **On additive generators of overlap functions**, [S.l.], Fuzzy Sets and Systems, v. 287, p. 76-96, 2016.

DIMURO, G. P.; BEDREGAL, B.; BUSTINCE, H.; JURIO, A.; BACZYNSKI, M.; MIS, K. **QL-operations and QL-implication functions constructed from tuples (O,G,N) and the generation of fuzzy subsethood and entropy measures**, [S.l.], International Journal of Approximate Reasoning, v. 82, p. 170-192, 2017.

DIMURO, G. P.; BEDREGAL, B.; BUSTINCE, H.; MESIAR, R.; ASIAN, M. J. **On Additive Generators of Grouping Functions**, [S.l.], Information Processing and Management of Uncertainty in Knowledge-Based Systems, v. 444, n. 252-261, 2014

DIMURO, G. P.; BEDREGAL, B. C.; LUCCA, G.; SANZ, J.A.; BUSTINCE, H.; MESIAR, R. **CMin-Integral: A Choquet-Like Aggregation Function Based on the Minimum t-Norm for Applications to Fuzzy Rule-Based Classification Systems**, [S.l.], Aggregation Functions in Theory and in Practice, p. 83-95, 2017.

DIMURO, G. P.; BEDREGAL, B.; SANTIAGO, R. H. N. **On  $(G, N)$ -implications derived from grouping functions**, [S.l.], Information Sciences, v. 279, p. 1-17, 2014.

DUBOIS, D.; OSTASIEWICZ, W.; PRADE, H. **Fuzzy Sets: History and Basic Notions**, Boston, Fundamentals of Fuzzy Sets, p. 21-124, 2000.

DUBOIS, D.; PRADE, H. **Fundamentals of Fuzzy Sets**, Boston, Kluwer, 2000.

ELKANO, M.; GALARA, M.; SANZ, J. A.; SCHIAVO, P. F.; PEREIRA, S. JR.; DIMURO, G. P.; BORGES, E. N.; BUSTINCE, H. **Consensus via penalty functions for decision making in ensembles in fuzzy rule-based classification systems**, [S.l.], Applied Soft Computing, 2017.

GAGOLEWSKI, M. **Spread measures and their relation to aggregation functions**, [S.l.], European Journal of Operational Research, v. 241, n. 2, p. 469-477, 2015.

GARCIA-JIMENEZ, S.; BUSTINCE, H.; HULLERMEIER, E.; MESIAR, R.; PAL, N. R.; PRADERA, A. **Overlap indices: Construction of and application to interpolative fuzzy systems**, [S.l.], IEEE Transactions on Fuzzy Systems, v. 23, n. 4, p. 1259-1273, 2015.

GRABISH, M.; MUROFUSHI, T.; SUGENO, M. **Fuzzy Measures and Integrals**, Heidelberg, Physica-Verlag, 2000.

ISHIBUCHI, H.; YAMAMOTO, T.; NAKASHIMA, T. **Hybridization of fuzzy GBML approaches for pattern classification problems**, [S.l.], IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), v. 35, n. 2, p. 359-365, 2005.

JURIO, A.; BUSTINCE, H.; PAGOLA, M.; PRADERA, A.; YAGER, R. R. **Some properties of overlap and grouping functions and their application to image thresholding**, [S.l.], Fuzzy Sets and Systems, v. 229, p. 69-90, 2013.

KURDILA, A. J.; ZABARANKIN, M. **Convex Functional Analysis**. Berlin, Springer, 2005.

LOURENZUTTI, R.; KROHLING, R. A.; REFORMAT, M. Z. **Choquet based TOPSIS and TODIM for dynamic and heterogeneous decision making with criteria interaction**, [S.l.], Information Sciences, v. 408, p. 41-69, 2017.

LOWEN, R. **Fuzzy set theory: basic concepts, techniques, and bibliography**, Dordrecht, Kluwer Academic Publisher, 1996.

LUCCA, G.; BUSTINCE, H.; SANZ, J.A.; DIMURO, G.P.; BEDREGAL, B.; ASIAIN, M.J.; ELKANO, M. **CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems**, Knowledge-Based Systems, v. 119, p. 32-43, 2016.

LUCCA, G.; BUSTINCE, H.; SANZ, J.A.; DIMURO, G.P.; BEDREGAL, B.; MESIAR, R.; KOLESAROVA, A.; OCHOA, G. **Pre-aggregation functions: Definition, Properties and Construction Methods**, IEEE Transactions on Fuzzy Systems, p. 260-272, 2016.

LUCCA, G.; DIMURO, G. P.; FERNANDEZ, J.; BUSTINCE, H.; BEDREGAL, B. C.; SANZ, J.A. **Improving the performance of fuzzy rule-based classification systems based on a non-averaging generalization of CC-integrals named CF1F2-integrals**, [S.l.], IEEE Transactions on Fuzzy Systems, 2018.

LUCCA, G.; SANZ, J.A.; DIMURO, G. P.; BEDREGAL, B. C.; BUSTINCE, H.; MESIAR, R. **CF-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems**, [S.l.], Information Sciences, v. 435, p. 94-110, 2018.

MAYOR, G.; TRILLAS, E. **On the representation of some aggregation functions**, Los Alamitos, Proceedings of IEEE International Symposium on Multiple-Valued Logic, IEEE, 1986, pp 111-114.

RIID, A.; RUSTERN, E. **Adaptability, interpretability and rule weights in fuzzy rule-based systems**, [S.l.], Information Sciences, v. 257, p. 301-312, 2014.

SANZ, J.; FERNANDEZ, A.; BUSTINCE, H.; HERRERA, F. **A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position**, International Journal of Approximate Reasoning, v. 52, p. 751-766, 2011.

SANZ, J.; FERNANDEZ, A.; BUSTINCE, H.; HERRERA, F. **IIVFDT: Ignorance functions based interval-valued fuzzy decision tree with genetic tuning**, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, v. 20, p. 1-30, 2012.

SANZ, J. A.; FERNANDEZ, A.; BUSTINCE, H.; HERRERA, F. **Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning**, [S.l.], Information Sciences, v. 180, n. 19, p. 3674-3685, 2010.

SUGENO, M. **Theory of Fuzzy Integrals and It's Applications**, Tokyo, Tese Ph.D., 1974.

WANG, Z.; KLIR, G. **Fuzzy Measure Theory**, New York, Plenum Press, 1992.

WILKIN, T.; BELIAKOV, G. **Weakly monotonic averaging functions**, [S.l.], International Journal of Intelligent Systems, v. 30, n. 2, p. 144-169, 2015.

YAGER, R.R. **On ordered weighted averaging aggregation operators in multicriteria decision making**, [S.l.], IEEE Trans. Syst. Man Cybern, v. 18, p. 183-190, 1988.

YAGER, R. R. **Toward a general theory of information aggregation**, [S.l.], Information Sciences, v.68, n. 3, p. 191-206, 1993.

YAGER, R. R.; RYBALOV, A. **Understanding the median as a fusion operator**, [S.l.], International Journal of General Systems, v. 26, n. 3, p. 239-263, 1997.

ZADEH, L. A. **Fuzzy sets as a basis for a theory of possibility**, [S.l.], Fuzzy Sets and Systems, v. 1, n. 1, p. 3-28, 1978.