



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
DEPARTAMENTO DE INFORMÁTICA E MATEMÁTICA APLICADA
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO



Generalizações da Integral de Choquet como Método de Combinação em Comitês de Classificadores

Thiago Vinícius Vieira Batista

Natal-RN
Março 2022

Thiago Vinícius Vieira Batista

Generalizações da Integral de Choquet como Método de Combinação em Comitês de Classificadores

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte como requisito para a obtenção do grau de Doutor em Ciência da Computação.

Linha de pesquisa:

Fundamentos da Computação

Orientador

Prof. Dr. Benjamín René Callejas Bedregal

Co-Orientador

Prof. Dr. Ronei Marcos de Moraes

PPGSC – PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

DIMAP – DEPARTAMENTO DE INFORMÁTICA E MATEMÁTICA APLICADA

CCET – CENTRO DE CIÊNCIAS EXATAS E DA TERRA

UFRN – UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

Natal-RN

Março 2022

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Batista, Thiago Vinícius Vieira.

Generalizações da integral de Choquet como método de combinação em comitês de classificadores / Thiago Vinícius Vieira Batista. - 2022.

118f.: il.

Tese (Doutorado) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Departamento de Informática e Matemática Aplicada, Programa de Pós-Graduação em Sistemas e Computação. Natal, 2022.

Orientador: Prof. Dr. Benjamín René Callejas Bedregal.

Coorientador: Prof. Dr. Ronei Marcos de Moraes.

1. Computação - Tese. 2. Comitês de classificadores - Tese. 3. Integral de Choquet - Tese. 4. Funções de pré-agregação - Tese. 5. Funções Overlap - Tese. 6. Funções de Quasi-Overlap - Tese. 7. Índices de validação - Tese. I. Bedregal, Benjamín René Callejas. II. Moraes, Ronei Marcos de. III. Título.

RN/UF/CCET

CDU 004

THIAGO VINÍCIUS VIEIRA BATISTA

“Generalizações da Integral de Choquet como Método de Combinação em Comitês de Classificadores”

Esta Tese foi julgada adequada para a obtenção do título de Doutor em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte.

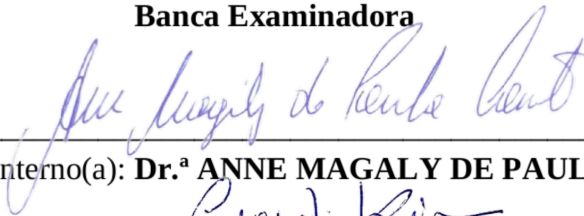


Prof. Dr. BENJAMÍN RENÉ CALLEJAS BEDREGAL
Presidente

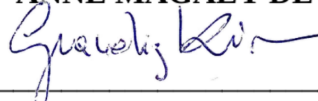


Prof.ª Dr.ª MONICA MAGALHÃES PEREIRA
Vice-coordenadora do PPgSC

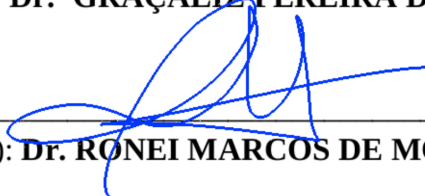
Banca Examinadora



Examinador(a) Interno(a): **Dr.ª ANNE MAGALY DE PAULA CANUTO**



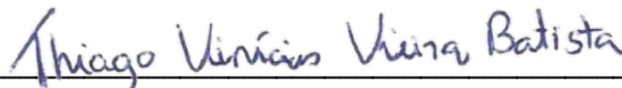
Examinador(a) Externo(a): **Dr.ª GRAÇALIZ PEREIRA DIMURO**



Examinador(a) Externo(a): **Dr. RONEI MARCOS DE MORAES**



Examinador(a) Externo(a): **Dr. RUI EDUARDO BRASILEIRO PAIVA**



Discente: **THIAGO VINÍCIUS VIEIRA BATISTA**

Natal, 04 de março de 2022

"O grande inimigo do conhecimento não é a ignorância, é a ilusão de ter conhecimento"

Stephen Hawking

Generalizações da Integral de Choquet como Método de Combinação em Comitês de Classificadores

Autor: Thiago Vinícius Vieira Batista

Orientador: Prof. Dr. Benjamín René Callejas Bedregal

Co-Orientador: Prof. Dr. Ronei Marcos de Moraes

RESUMO

Comitê de classificadores é um modelo dentro da aprendizagem de máquina que consiste em uma coleção de classificadores que processam uma mesma informação e a saída destes é combinada de alguma maneira. O processo de classificação geralmente se dá em duas etapas: a etapa de classificação e a etapa de combinação. Na etapa de classificação, cada classificador processa a informação e fornece sua opinião sobre a classe à qual o elemento pertence, na etapa de combinação, a opinião de todos os classificadores são combinadas, fornecendo uma única saída. Apesar da etapa de combinação ser de grande importância, muitos dos trabalhos encontrados na literatura focam principalmente na etapa de classificação. Neste trabalho serão propostas generalizações da integral de Choquet para uso como método de combinação em comitês de classificadores. A ideia principal é permitir uma maior liberdade de escolha de funções para compor a integral, abrindo possibilidades de otimização e utilização de funções adequadas aos dados. Além disso, é proposta uma nova noção de monotonicidade parcial e conseqüentemente uma alternativa à noção de funções de pré-agregação. Resultados preliminares obtidos pelas generalizações da integral de Choquet propostas mostraram que elas foram capazes de obter bons resultados, tendo tido desempenho superior a métodos conhecidos da literatura como o XGBoost, Bagging, entre outros. Além disso, as integrais que envolveram as novas funções de agregação propostas tiveram um bom desempenho quando comparadas com a performance ao se utilizar outras classes de funções, como as Cópulas e os Overlaps.

Palavras-chave: Comitês de Classificadores; Integral de Choquet; Funções de Pré-Agregação; Funções Overlap; Funções de Quasi-Overlap; Índices de Validação

Generalizations of the Choquet Integral as a Combination method in Ensemble of Classifiers

Author: Thiago Vinícius Vieira Batista

Supervisor: Dr. Benjamín René Callejas Bedregal

Co supervisor: Dr. Ronei Marcos de Moraes

ABSTRACT

Ensembles of classifiers is an method in machine learning that consists in a collection of classifiers that process the same information and their output is combined in some manner. The process of classification is done in two main steps: the classification step and the combination step. In the classification step, each classifier processes the information and provides an output, in the combination step, the output of every classifier is combined, providing a single output. Although the combination step is extremely important, most works focus mostly on the classification step. Therefore, in this work, generalizations of the Choquet Integral will be proposed to be used as a combination method in ensembles of classifiers. The main idea is to allow a greater freedom of choice for functions in the integral, opening possibilities for optimization and using functions adequate to the data. Furthermore, a new notion of partial monotonicity is proposed, and consequently an alternative to the notion of pre-aggregation functions. Preliminary results that were obtained by the generalizations of the Choquet integral in the ensemble showed that they were capable of obtaining good results, having a superior performance to known methods in literature such as XGBoost, Bagging, among others. Furthermore, the generalizations that used the proposed aggregation functions had good performance when compared to other classes of functions, such as Copulas and Overlaps.

Keywords: Ensembles of Classifiers, Choquet Integral, Pre-Aggregation Functions

Lista de figuras

1	Exemplo de funcionamento do k - NN com $k = 3$	p. 23
2	Exemplo de árvore de decisão. Adaptado de (MITCHELL, 1997)	p. 24
3	Configurações de uma rede neural. Adaptados de (HAYKIN, 2004)(a) e (RUSSELL; NORVIG, 2002) (b)	p. 25
4	Arquitetura de um comitê de classificadores	p. 27
5	Exemplo de diagrama CD comparando a performance de 4 classificadores	p. 33
6	Comparação da performance do método ao variar a quantidade de classificadores presentes no comitê	p. 81
7	Diagrama CD para comparação dos métodos que utilizaram a integral CQO	p. 82
8	Diagrama CD para comparação entre os métodos externos e a melhor configuração da integral CQO	p. 84
9	Diagrama CD para comparação entre os métodos externos e a melhor configuração da CQO-Integral e AC-Integral	p. 90
10	Comparação da acurácia do comitê pela quantidade de classificadores presentes no comitê	p. 92

Lista de tabelas

1	Matriz de confusão para o caso binário	p. 29
2	Tabela com exemplos de diferentes tipos de cópulas, <i>overlaps</i> e <i>quasi-overlaps</i>	p. 37
3	Definição de alguns <i>pseudo-quasi overlaps</i> e outras funções que serão utilizadas neste trabalho	p. 57
4	Matriz V com o resultado das classificações	p. 74
5	Descrição das bases de dados utilizadas neste trabalho	p. 78
6	Resultados do experimento realizado com a CQO-Integral e a medida <i>fuzzy</i> uniforme	p. 82
7	Resultados da aplicação de outros métodos disponíveis na literatura	p. 83
8	Resultados da aplicação de outros métodos disponíveis na literatura	p. 85
9	Médias da acurácia dos resultados das classificações com a AC-Integral para cada par de combinações utilizando a medida uniforme	p. 85
10	Resultados do experimento realizado com a AC-Integral e a medida <i>fuzzy</i> uniforme	p. 86
11	Médias da acurácia dos resultados das classificações com a AC-Integral para cada par de combinações utilizando a medida <i>fuzzy</i> ponderada	p. 87
12	Ganho de desempenho das combinações após a utilização da medida <i>fuzzy</i> ponderada	p. 88
13	Resultados da aplicação de outros métodos disponíveis na literatura	p. 89
14	Diferença entre o desempenho da melhor configuração da AC-Integral contra outros métodos	p. 89
15	Resultados da aplicação de outros métodos disponíveis na literatura	p. 91

16	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^1 com os outros métodos	p. 103
17	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^2 com os outros métodos	p. 104
18	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^3 com os outros métodos	p. 104
19	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^4 com os outros métodos	p. 105
20	Resultados da aplicação da AC-Integral utilizando a função O_Q^1 com os outros métodos	p. 105
21	Resultados da aplicação da AC-Integral utilizando a função O^1 com os outros métodos	p. 106
22	Resultados da aplicação da AC-Integral utilizando a função C^1 com os outros métodos	p. 107
23	Resultados da aplicação da AC-Integral utilizando a função O^2 com os outros métodos	p. 108
24	Resultados da aplicação da AC-Integral utilizando a função O_Q^2 com os outros métodos	p. 109
25	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^1 com os outros métodos	p. 110
26	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^2 com os outros métodos	p. 111
27	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^3 com os outros métodos	p. 111
28	Resultados da aplicação da AC-Integral utilizando a função O_{PO}^4 com os outros métodos	p. 112
29	Resultados da aplicação da AC-Integral utilizando a função O_Q^1 com os outros métodos	p. 112
30	Resultados da aplicação da AC-Integral utilizando a função O^1 com os outros métodos	p. 113

31	Resultados da aplicação da AC-Integral utilizando a função C^1 com os outros métodos	p. 114
32	Resultados da aplicação da AC-Integral utilizando a função O^2 com os outros métodos	p. 115
33	Resultados da aplicação da AC-Integral utilizando a função O_Q^2 com os outros métodos	p. 116

Lista de abreviaturas e siglas

NB – Naive Bayes

k-NN – *k-Nearest Neighbors*

RNA – Redes Neurais Articiais

AD – Árvores de Decisão

VP – Verdadeiros Positivos

VN – Verdadeiros Negativos

FP – Falsos Positivos

FN – Falsos Negativos

SVM – *Support Vector Machines*

GM – *General Mixture*

CQO-Integral– Integral de Choquet Baseada em Quasi-Overlaps

AC-Integral – Integral de Choquet Absoluta

RF – Random Forests

ADA – AdaBoost

BAG – Bagging

XGB – XGBoost

META – META-DES

Lista de Algoritmos

1	Pseudo-código para classificação de elementos o	p. 71
2	Pseudo-código para o modelo proposto	p. 72
3	Pseudo-código para cálculo da AC-Integral	p. 73
4	Pseudo-código para o modelo proposto para definição de pesos	p. 76

Sumário

1	Introdução	p. 16
1.1	Motivação	p. 17
1.2	Objetivos	p. 18
1.3	Organização do Trabalho	p. 19
2	Fundamentação Teórica	p. 20
2.1	Aprendizagem de Máquina	p. 20
2.1.1	Aprendizagem Supervisionada	p. 21
2.1.2	Algoritmos de Aprendizagem Supervisionada	p. 22
2.1.2.1	Naive Bayes	p. 22
2.1.2.2	<i>k-Nearest Neighbors</i>	p. 23
2.1.2.3	Árvores de Decisão	p. 24
2.1.2.4	Redes Neurais Artificiais	p. 25
2.2	Comitês de Classificadores	p. 26
2.3	Métricas de Performance para Classificadores	p. 28
2.3.1	Matriz de Confusão e Acurácia	p. 28
2.3.2	Precisão e Recall	p. 29
2.3.3	Medida F_β	p. 30
2.3.4	Kappa de Cohen	p. 30
2.3.5	Métricas sobre Problemas Multi-Classe	p. 30
2.4	Testes Estatísticos	p. 31
2.5	Conceitos Matemáticos	p. 33

2.5.1	Medidas Fuzzy	p. 36
2.5.2	Integral de Choquet	p. 39
2.6	Considerações Finais	p. 41
3	Trabalhos Correlatos	p. 42
3.1	Generalizações da Integral de Choquet	p. 42
3.2	Comitês de Classificadores	p. 43
3.2.1	Métodos de Criação de Comitês de Classificadores	p. 43
3.2.2	Métodos de Combinação baseados em Fusão	p. 45
3.2.2.1	Combinação de fusão <i>hard</i>	p. 45
3.2.2.2	Combinação de fusão <i>soft</i>	p. 46
3.2.2.3	Métodos de Combinação com Integrais de Choquet	p. 46
3.3	Considerações Finais	p. 47
4	Contribuições Teóricas	p. 48
4.1	<i>Pseudo Overlaps</i>	p. 48
4.2	<i>Pseudo-Quasi Overlaps</i>	p. 51
4.3	Funções de (\cdot) -pré-agregação	p. 57
4.4	Novas Generalizações da Integral de Choquet	p. 58
4.4.1	Integral de Choquet Utilizando <i>Quasi-Overlaps</i>	p. 58
4.4.2	Integral de Choquet Utilizando <i>Pseudo Overlaps</i>	p. 62
4.4.3	Integral de Choquet Absoluta	p. 64
4.5	Considerações Finais	p. 70
5	Metodologia Proposta e Configuração Experimental	p. 71
5.1	Arquitetura do Comitê de Classificadores Proposta	p. 71
5.2	Exemplo de Funcionamento da Arquitetura Proposta	p. 73
5.3	Procedimento para Definição de Pesos	p. 75

5.4	Configuração dos Experimentos	p. 77
5.4.1	Bases de Dados	p. 77
5.4.2	Configurações de Parâmetro e do Comitê	p. 78
5.5	Considerações Finais	p. 79
6	Resultados	p. 80
6.1	Determinação do número de classificadores no comitê	p. 80
6.2	Experimentos com a CQO-Integral	p. 81
6.3	Experimentos com a AC-Integral	p. 84
6.4	Discussão dos resultados	p. 90
6.5	Considerações Finais	p. 92
7	Considerações Finais	p. 93
7.1	Considerações	p. 93
7.2	Trabalhos Futuros	p. 95
7.3	Produções Bibliográficas	p. 95
	Referências	p. 96
	Apêndice A – Resultados das classificações para cada par de combinações com a AC-Integral utilizando a medida uniforme	p. 103
	Apêndice B – Resultados das classificações para cada par de combinações com a AC-Integral utilizando a medida ponderada	p. 110
	Apêndice C – Descrição das Bases de Dados	p. 117

1 Introdução

Classificadores são modelos computacionais que são capazes de mapear elementos em um conjunto de categorias discretas chamadas classes, baseado em um conjunto de variáveis destes elementos. Esses classificadores pertencem a uma área do conhecimento chamada aprendizagem de máquina, e tem sido aplicados em muitas áreas de conhecimento que envolvem algum tipo de tomada de decisão.

Uma estratégia comumente utilizada ao se trabalhar com classificação é selecionar um conjunto de classificadores, formando o que se conhece como comitê de classificadores. Esse comitê funciona de maneira semelhante a um comitê na vida real, com cada classificador processando um determinado elemento e apresentando uma decisão sobre a classificação deste, e então essas decisões são combinadas em uma única resposta, normalmente levando em consideração a classificação realizada por todos os classificadores (MOHANDES; DERICHE; ALIYU, 2018).

A ideia por trás de se utilizar vários classificadores é, através da diversidade, melhorar o desempenho do sistema de classificação em geral (KUNCHEVA; WHITAKER, 2003). Em elementos onde determinados classificadores tem performance reduzida, a ideia é que os outros compensem essa deficiência. Além disso, essa abordagem tende a reduzir problemas comuns em classificação como os mínimos locais e escolhas incorretas de classificadores.

Em comitês de classificadores, talvez a parte mais importante seja a combinação dos resultados dos classificadores. Nos tempos atuais onde trabalhar com um grande volume de dados se tornou uma atividade recorrente e importante, além de ter boa precisão, métodos de classificação também precisam ter boa performance computacional. Portanto, uma estratégia de combinação deve ser tanto computacionalmente eficiente quanto ter boa performance.

Existem diferentes abordagens para a etapa de combinação. Em alguns sistemas, alguns dos classificadores tem seus resultados descartados (YANG, 2011) baseado em alguma métrica sobre sua performance, esta abordagem é conhecida como **seleção**. Em outros,

uma função matemática ou uma votação agrega os resultados dos classificadores, sendo esta abordagem conhecida como **fusão**. Em ambas abordagens, existem variações que introduzem etapas adicionais. Por exemplo, existem estratégias que introduzem uma etapa adicional de treinamento para determinar quais classificadores irão contribuir mais com a resposta final (LIU; DU; WU, 2019).

Uma abordagem interessante para as abordagens de fusão são as funções de agregação. Elas são uma família de funções matemáticas que combinam um conjunto de entrada em um único valor (BELIAKOV et al., 2007) e usualmente são funções razoavelmente simples e que não consomem muitos recursos computacionais. Em Bustince et al. (2016), a ordem atrelada à noção de funções de agregação foi relaxada dando origem ao conceito de pré-agregação. Uma classe especial de funções de pré-agregação que vem recebendo atenção nos últimos anos são as variantes ou generalizações das integrais de Choquet. Essas funções tem apresentado excelente resultados em diferente áreas (DIMURO et al., 2020a; BUSTINCE et al., 2021) como avaliação de estudantes (SHIEH; WU; LIU, 2009), reconhecimento facial (KARCZMAREK; KIERSZTYN; PEDRYCZ, 2018) e previsão do tempo (ABDEL-NASSER; MAHMOUD; LEHTONEN, 2020).

Neste trabalho, serão propostas abordagens que visam melhorar a classificação em comitês de classificadores, através da definição de extensões das integrais de Choquet, que irão atuar na etapa de combinação da saída dos classificadores, utilizando a abordagem de fusão. Também serão propostas generalizações das funções de *overlap* (BUSTINCE et al., 2010a) através do afrouxamento das restrições de comutatividade e continuidade. Essa família de funções tem sido explorada em trabalhos como o de Paiva et. al (2021b), e tem apresentado bons resultados em aplicações de classificação. Uma contribuição final deste trabalho é a proposição de uma nova forma de relaxar a monotonicidade das funções de agregação.

1.1 Motivação

Existe uma grande quantidade de métodos de comitês de classificadores na literatura, visto que esse método tem sido amplamente utilizado em diversas áreas do conhecimento devido a sua capacidade de generalização, redução do risco de *overfitting* e maior chance de encontrar a solução ótima do problema através da combinação de diversos classificadores. Além disso, ao se escolher múltiplos classificadores, o risco de escolher um único classificador inadequado diminui (MOYANO et al., 2018).

Devido a suas vantagens, esses métodos tem sido aplicados em áreas como detecção de doenças em imagens médicas (DEY et al., 2021; ATALLAH; AL-MOUSA, 2019), no mercado financeiro (CAO, 2012), controle de qualidade da água (KADKHODAEI; MOGHADAM; DEHGHAN, 2020) e sistemas de recomendação (SCHCLAR et al., 2009). No entanto, grande parte dos trabalhos não trata da etapa de combinação de classificadores, utilizando métodos simples como a votação por maioria ou métodos baseados em uma média ponderada dos resultados. Além disso, esses métodos de combinação funcionam em determinadas condições, que nem sempre acontecem em uma aplicação real.

Desta forma, a principal motivação deste trabalho é propor um novo método de combinação que consiga melhorar o desempenho de comitês de classificadores. Neste contexto, a escolha pela integral de Choquet se deu pelos bons resultados que este método vem obtendo em outros trabalhos. Embora alguns trabalhos (LI; WANG; CHEN, 2015; PACHECO; KROHLING, 2018; CAO, 2012) já utilizem a integral de Choquet no contexto de comitês de classificadores, eles a utilizam em sua versão clássica, focando principalmente na mudança da medida *fuzzy*, que é um dos elementos da integral.

1.2 Objetivos

O objetivo geral deste trabalho é a definição de novas generalizações da Integral de Choquet para uso como método de combinação em comitês de classificadores. Para tal, foram definidos os seguintes objetivos específicos:

- Definir duas novas classes de funções de agregação chamadas *pseudo overlaps* e *pseudo-quasi overlaps*, que são generalização dos *overlaps* definidos a partir do relaxamento das condições de continuidade, comutatividade ou ambas;
- Definir generalizações da integral de Choquet baseada em *quasi-overlaps* e *pseudo-overlaps*, que por sua vez são uma generalizações de funções de *overlap*;
- Apresentar a Integral de Choquet Absoluta, uma generalização das integrais de Choquet para ser utilizada em conjunto com funções de agregação;
- Realizar experimentos com as generalizações da integral de Choquet, comparando estes métodos entre si e com alguns dos principais métodos de comitês de classificadores;
- Aplicar os *pseudo-quasi overlaps* no contexto das generalizações da integral de Choquet e verificar se houve um ganho de performance;

- Validar os resultados através de testes estatísticos.

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte maneira:

- No capítulo 2 serão definidas os principais conceitos necessários para o entendimento das contribuições desta tese.
- No capítulo 3 serão apresentados alguns trabalhos que tiveram objetivos ou metodologias similares a esta tese.
- No capítulo 4 serão apresentadas as contribuições desta tese, principalmente aquelas que são consideradas teóricas
- No capítulo 5, as definições metodológicas e experimentais, além da abordagem proposta para a configuração do comitê de classificadores
- No capítulo 6 serão apresentados os resultados obtidos a partir da utilização das generalizações das Integrais de Choquet propostas no contexto de classificadores. Será também realizada uma análise e discussão dos resultados, além de comparação com outros métodos similares
- Finalmente, no capítulo 7 serão feitas algumas considerações finais com relação aos resultados obtidos e serão apresentadas as produções resultantes desta tese.

2 Fundamentação Teórica

Este capítulo apresentará conceitos que serão importantes para o desenvolvimento e compreensão desta tese. Primeiramente, trataremos do conceito de aprendizagem de máquina, focando na área de interesse do trabalho, que é a aprendizagem supervisionada utilizando comitês de classificadores. Posteriormente, serão lembrados conceitos matemáticos presentes na literatura: Funções de Agregação, Cópulas, Overlaps, Quasi-Overlaps, Medidas *Fuzzy* e Integrais de Choquet.

2.1 Aprendizagem de Máquina

Aprendizagem de máquina é uma área do conhecimento nascida da ciência da computação que tem como objetivo tornar o computador capaz de *aprender* a tomar decisões relativas a uma determinada tarefa. Nesta definição, o conceito de aprendizado refere-se à capacidade do computador de melhorar sua performance com respeito a alguma tarefa, após adquirir conhecimento e experiência sobre a mesma (MITCHELL, 1997).

De forma mais prática, podemos definir a aprendizagem de máquina como a construção de modelos computacionais capazes de adquirir conhecimento e tomar decisões sobre futuros problemas a que sejam apresentados. Em geral, o aprendizado se dá na chamada etapa de treinamento, onde é apresentado ao modelo um conjunto de dados referente ao problema. Então, a partir deste conjunto, o modelo é capaz de inferir regras e parâmetros que são usados para classificar novos exemplos.

Dependendo da forma como o conjunto de dados de treino é apresentado ao modelo, é possível dividir os modelos de aprendizagem de máquina em três principais categorias (RUSSELL; NORVIG, 2002):

- Quando os dados apresentados vem acompanhados de uma indicação de qual sua correta classificação, dizemos que é um modelo de **aprendizagem supervisionada**;

- Quando não há qualquer indicação sobre a classificação dos dados sendo apresentados, dizemos que temos um modelo de **aprendizagem não supervisionada**;
- Finalmente, quando a única informação é o fornecimento de um *feedback* qualitativo sobre uma determinada decisão quando ela for tomada, mas sem informar qual seria a melhor decisão, temos um modelo de **aprendizagem por reforço**.

Nesta tese, serão utilizados apenas modelos de aprendizagem supervisionada, portanto, a teoria sobre as outras duas categorias de modelos não serão aprofundadas. No entanto, de forma geral, modelos de aprendizagem não supervisionada são importantes para identificar padrões em conjuntos de dados, aprendendo como os elementos destes conjuntos se relacionam. Já modelos de aprendizagem por reforço são comumente utilizados onde é muito difícil ou impossível prover um conjunto de dados que preveja todas as possíveis ações e seus resultados, como por exemplo, em jogos (RUSSELL; NORVIG, 2002).

2.1.1 Aprendizagem Supervisionada

O modelo de aprendizagem supervisionada envolve aprender a relação entre elementos e um valor que indica sua correta classificação e utilizar esta relação para prever novos elementos (CUNNINGHAM; CORD; DELANY, 2008). Isso ocorre na **etapa de treinamento**, onde o classificador aprende a realizar esse mapeamento para ser capaz de receber os novos elementos. Quando se deseja aferir a efetividade do classificador, é comum realizar a chamada **etapa de teste**, onde é apresentado ao classificador elementos cuja classificação já é conhecida, e então verifica-se se o classificador atribuiu a classe correta a estes elementos.

Nesse contexto, um elemento referente a um conjunto de dados é um vetor $\mathbf{X} = (X_1, \dots, X_k)$ de k elementos, chamados de **atributos** do elemento X . Cada um desses atributos X_i possui seu próprio domínio D_i . O conjunto de todos os possíveis elementos (ou domínio do problema) é $\Omega = D_1 \times \dots \times D_k$. **Um conjunto de dados** D , portanto, é um subconjunto de Ω (RISH et al., 2001).

Para cada $\mathbf{X} \in \Omega$ podemos associar uma categoria $y \in \{c_1, \dots, c_p\}$, que também é conhecida como **classe**. Esse valor pode ser fornecido previamente (na etapa de treinamento) ou ser o elemento a ser inferido pelo classificador, levando em consideração as informações e regras que conseguiu extrair na etapa de treinamento. No contexto de aprendizagem supervisionada um classificador é, portanto, uma mapeamento $C : \Omega \rightarrow \{c_1, \dots, c_p\}$.

Modelos de aprendizagem supervisionada vem sendo aplicados com muito êxito em diferentes áreas como bolsa de valores (NABIPOUR et al., 2020), reconhecimento de gestos (BATISTA et al., 2019) e classificação de imagens (HONG et al., 2021). São usados comumente para prever novas situações baseadas em eventos que já ocorreram. Nos últimos anos, intensos estudos na área acarretaram o surgimento de diversos algoritmos que têm diferentes vantagens e desvantagens.

2.1.2 Algoritmos de Aprendizagem Supervisionada

Nesta seção, serão descritos alguns dos principais e mais conhecidos modelos de aprendizagem supervisionada utilizados, que são aplicados nesta tese posteriormente. Aqui apresentaremos os seguintes modelos: Naive Bayes – NB , *k-Nearest Neighbors* – k-NN , Redes Neurais Artificiais – RNA e Árvores de Decisão – AD .

2.1.2.1 Naive Bayes

O Naive Bayes é uma família de classificadores probabilísticos. Este modelo é baseado no Teorema de Bayes, descrito a seguir, utilizando uma forte suposição de independência. Seja c_i uma classe e \mathbf{X} um elemento de entrada:

$$P(c_i|\mathbf{X}) = \frac{P(\mathbf{X}|c_i)P(c_i)}{P(\mathbf{X})} \quad (2.1)$$

A Equação (2.1) representa o Teorema de Bayes. Usando a hipótese que todas os atributos são condicionalmente independentes entre si dada a classe, pode-se reescrever a Equação (2.1) da seguinte forma:

$$P(c_i|\mathbf{X}) = P(c_i) \prod_{j=1}^k P(x_j|c_i) \quad (2.2)$$

Onde $P(c_i)$ é a probabilidade *a priori* para a classe c_i e $P(x_j|c_i)$ é a probabilidade condicional de x_j dado que a classe é c_i . O processo de decisão deste modelo, então, é calcular, para cada $c_i \in \{c_1, \dots, c_p\}$ a probabilidade $P(c_i|\mathbf{X})$ descrita na Equação (2.2) e atribuir ao elemento \mathbf{X} a classe que obteve o maior valor (MORAES; SOARES; MACHADO, 2020).

A grande vantagem deste modelo é a simplicidade, pois para realizar a classificação, basta calcular as probabilidades previamente, que possuem um número finito graças à

suposição de independência. Além disso, apesar da simplicidade, este modelo gera bons resultados (PHAM et al., 2016). No entanto, a desvantagem é que em bases de dados em que existe um alto grau de correlação entre as variáveis, este modelo tende a ter uma queda de performance.

2.1.2.2 *k-Nearest Neighbors*

O algoritmo *k-Nearest Neighbors* (ou *k-vizinhos mais próximos*, em português) é um algoritmo de aprendizagem supervisionada em que os elementos são classificados baseado na classe dos k vizinhos mais próximos. Estes são calculados baseados em métricas de distância, como a distância Euclidiana. A forma de como usar a classe dos vizinhos selecionados também pode variar, sendo a mais comum a votação por maioria (CUNNINGHAM; DELANY, 2007).

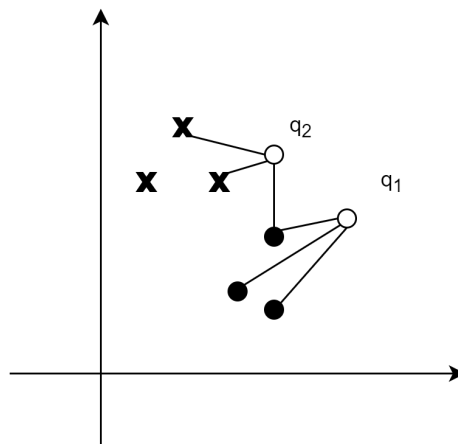


Figura 1: Exemplo de funcionamento do k -NN com $k = 3$

Na Figura 1, pode-se verificar o exemplo do funcionamento do k -NN com $k = 3$. No caso do elemento q_2 , por exemplo, vemos que após o cálculo da distância, 2 dos seus três vizinhos tem a classe "x", portanto, num cenário em que se usa a votação por maioria, esta seria a classe atribuída a q_2 . Caso tivéssemos um k par e ocorresse uma situação de empate, normalmente atribui-se aleatoriamente uma classe entre as que tiveram a menor distância.

Este é, talvez, o mais simples método de classificação supervisionada. Foi extensivamente estudado e muitas variações do seu funcionamento foram propostas. O motivo para a longevidade deste método se dá principalmente pela simplicidade, aliado a bons resultados que são obtidos. No entanto, esse método requer uma estimação custosa para estimar um valor de k (MAKKAR et al., 2017), e a complexidade aumenta proporcionalmente ao

valor de k escolhido.

2.1.2.3 Árvores de Decisão

Árvores de decisão é um outro algoritmo de classificação que constrói uma estrutura de árvore para representar o processo de tomada de decisão, onde os nós da árvore representam os atributos, as arestas que saem dos nós representam todas os possíveis valores do atributo no nó de onde esta aresta partiu, e finalmente as folhas representam as decisões.

O processo de tomada de decisão neste modelo é feito de forma intuitiva, onde, dado um elemento X , deve-se percorrer a árvore partindo da raiz até chegar em uma folha, que representa a decisão final. Um exemplo desta árvore está representado na Figura 2, que esquematiza o processo de tomada de decisão para ir ou não jogar tênis.

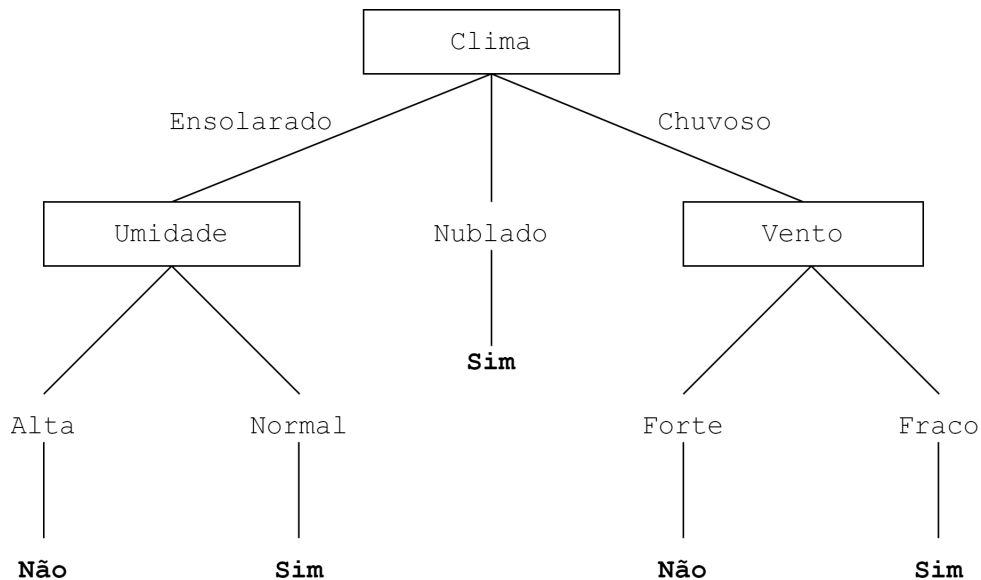


Figura 2: Exemplo de árvore de decisão. Adaptado de (MITCHELL, 1997)

Seja $\Omega = D_{clima} \times D_{umidade} \times D_{vento}$. Utilizando a árvore da Figura 2 como exemplo e considerando um elemento $\mathbf{X} = (\text{Ensolarado}, \text{Alta}, \text{Fraco})$, teríamos a decisão **Não**. Uma das grandes vantagens deste método, como pode-se perceber, é sua representatividade visual, visto que o processo de tomada de decisão neste modelo é intuitivo, ao contrário de outros modelos como o Naive Bayes e as Redes Neurais. Além disso, definida a estrutura da árvore, o processo de classificação se dá de maneira relativamente rápida, necessitando de poucos recursos computacionais.

O processo de construção deste modelo varia conforme a implementação, visto que da mesma forma que o k-NN, existem muitas abordagens na literatura para construção.

Porém, em geral, essas árvores são construída da raiz até as folhas, de maneira *top-down* utilizando uma abordagem gulosa.

Inicialmente, é verificado qual atributo irá contribuir mais para a classificação, e então, este atributo é colocado na raiz da árvore, com todos os possíveis valores deste atributo sendo colocados como arestas na árvore partindo deste nó. No nível seguinte, é feita a mesma verificação com os atributos restantes e realizando o mesmo processo, até que se gere um nó folha, que é a decisão. A verificação do atributo que mais irá contribuir para a classificação pode ser feita, por exemplo, através do cálculo do ganho de informação que é baseado no conceito de entropia dos atributos (MITCHELL, 1997).

2.1.2.4 Redes Neurais Artificiais

As redes neurais são uma família de modelos de aprendizagem supervisionada inspiradas na organização do cérebro humano, composta por uma rede de neurônios que se comunicam e são capazes de aprender, analisando padrões de atividades, com a informação aprendida armazenada como pesos nas conexões entre os neurônios (MEDSKER, 2012). Uma rede neural é, portanto, composta por um conjunto de nós (representando os neurônios), que são conectados por arcos (representando as sinapses), onde a cada arco é associado um peso, que é utilizado para armazenar o conhecimento da rede.

Naturalmente, existem muitas arquiteturas e diferentes abordagens para a construção dessas redes. Nesta tese, focaremos em uma abordagem clássica, visto que apenas utilizaremos estes métodos futuramente. Desta forma, utilizaremos a arquitetura *Multilayer Perceptron*.

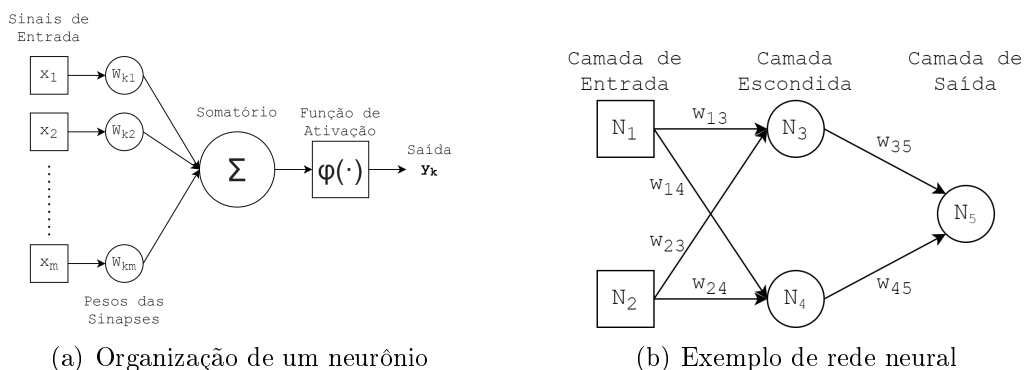


Figura 3: Configurações de uma rede neural. Adaptados de (HAYKIN, 2004)(a) e (RUSSELL; NORVIG, 2002) (b)

A arquitetura deste modelo consiste em três camadas principais: a camada de entrada, as camadas escondidas e a camada de saída, como pode ser visto na rede neural de

exemplo na Figura 3(b). A camada de entrada consiste de um número n de neurônios que representam os sinais de entrada. Os neurônios nesta camada normalmente realizam pouco processamento sobre os dados e apenas repassam o sinal à camada seguinte (RUSSELL; NORVIG, 2002).

As camadas escondidas são um conjunto de camadas intermediárias entre a camada de entrada e a de saída. Nela, os neurônios normalmente obedecem a estrutura da Figura 3(a), que consiste de três elementos principais (HAYKIN, 2004):

1. Um conjunto de arestas ou sinapses que chegam ao neurônio, cada uma com um peso w_{kj} , referente ao peso da aresta que conecta o neurônio k ao neurônio j multiplicado pelo sinal x_k sendo passado pela aresta.
2. Um somatório que soma os valores de entrada após terem sido multiplicados pelos pesos das arestas, formando um combinador linear.
3. Uma função de ativação (φ) que limita a amplitude do sinal que vai ser propagado pelo neurônio. Uma função que é muito utilizada nestes sistemas é a função sigmóide.

Finalmente, a camada de saída é o resultado final do processamento da rede, e seus neurônios também obedecem a estrutura da figura 3(a). Neste modelo, o treinamento da rede é realizado através do algoritmo de retropropagação do erro, que utiliza regressão para minimizar uma função de erro, alterando os pesos para que o sinal de saída seja o mais próximo possível dos valores corretos da classificação (AGGARWAL et al., 2018).

2.2 Comitês de Classificadores

Comitê de classificadores é uma abordagem que combina múltiplos classificadores para resolver o mesmo problema. Nestes sistemas, o mesmo conjunto de dados é apresentado para um conjunto de classificadores, e então a resposta destes é combinada para gerar uma única decisão (ZHOU, 2019). Dependendo do tipo dos classificadores presentes no comitê, pode-se classificar estes em dois tipos: **heterogêneos** e **homogêneos**. Em comitês homogêneos, geralmente são usados classificadores do mesmo tipo, já em comitês heterogêneos, diferentes tipos classificadores podem ser utilizados.

A Figura 4 representa a arquitetura básica desta abordagem. Conforme apresentado na figura, podemos dividir este modelo em duas camadas principais: a camada de classi-

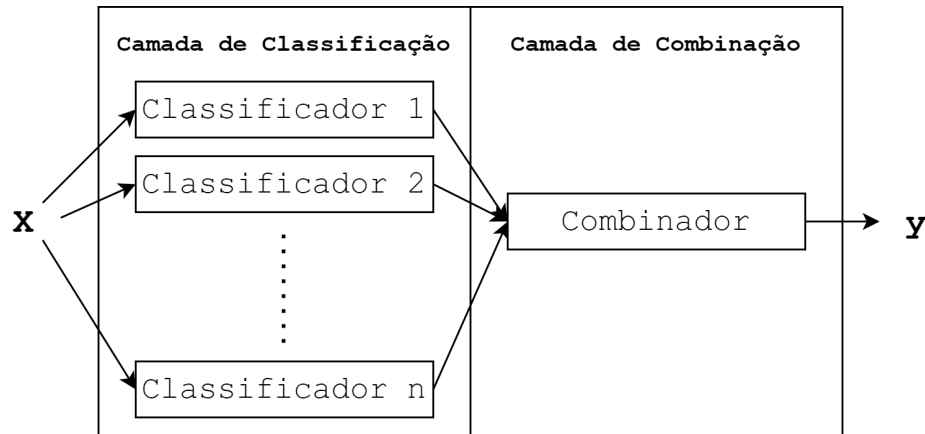


Figura 4: Arquitetura de um comitê de classificadores

ficação e a camada de combinação. Na camada de classificação, o mesmo problema \mathbf{X} é apresentado para o comitê e processado individualmente por cada classificador.

A camada seguinte é a camada de combinação, que será o foco nesta tese. Nesta, as saídas dos n classificadores são combinadas em um único valor. Existem diferentes estratégias para realizar essa combinação mas elas podem ser divididas em dois principais grupos:

- Na abordagem **seletiva**, a saída de determinados classificadores é descartada, baseado em alguma métrica de performance. Em (KRAWCZYK; WOZNIAK, 2016), uma função Gaussiana foi usada para atribuir valores para cada classificador baseado em sua performance, e então classificadores que tivessem um valor abaixo de um certo limite eram descartados no cálculo do resultado final. Outros exemplos de aplicações desta estratégia são algoritmos evolucionários (KRAWCZYK; WOZNIAK, 2018) e árvores de decisão (DING et al., 2018).
- Na abordagem de **fusão**, a saída de todos os classificadores é levada em consideração. Esta abordagem inclui abordagens simples como votação por maioria, utilizada no trabalho de (KODOVSKY; FRIDRICH; HOLUB, 2011). Uma outra abordagem possível é a atribuição de pesos para os classificadores, determinando o peso de cada classificador no resultado final. Idealmente, classificadores com melhor performance receberão um peso maior. Estes pesos podem ser estáticos, ou seja, atribuídos de antemão, ou dinâmicos, quando são calculados em tempo de execução. Por exemplo, Costa et. al 2018 utilizou uma família de funções de *General Mixture* (GM) para combinar a saída dos classificadores. Os pesos neste métodos foram calculados baseado na distância de cada saída para um valor de referência. Outros trabalhos ainda

podem realizar uma etapa de treinamento adicional para geração destes pesos (LIU; DU; WU, 2019).

De acordo com Dietterich (2000), existem três principais razões pelas quais comitês de classificadores são uma boa abordagem para classificação, geralmente fornecendo melhores resultados que quando utiliza-se os classificadores individualmente:

1. **Estatística:** Quando o conjunto de dados em que o algoritmo está sendo treinado é muito menor que o espaço das hipóteses, o classificador pode encontrar várias hipóteses que dão a mesma acurácia e que não são boas para prever casos futuros. Ao combinar os classificadores, este risco é reduzido pois estes combinam seus votos e reduzem o risco de tal situação acontecer.
2. **Computacional:** Visto que muitos classificadores realizam buscas locais no espaço, e estes podem ficar presos em mínimos locais que não são a solução ótima do problema. Portanto, utilizar classificadores diferentes ou o mesmo classificador começando em pontos diferentes do espaço, esta situação pode ser mitigada.
3. **Representacional:** Para um problema de classificação, às vezes a melhor hipótese não pode ser representada por nenhuma hipótese no espaço. Através da combinação de classificadores, é possível expandir o espaço de funções representáveis, e a melhor hipótese pode ser aproximada com mais precisão.

De acordo com Zhou (2019), geralmente o custo computacional de construir um comitê de classificadores não é muito maior do que utilizar um único classificador, pois os processos de calibração que precisam ser realizados tem custo similar ao de se criar um comitê. Além disso, as técnicas utilizadas para combinação geralmente são simples e não acarretam em um grande aumento de recursos computacionais.

2.3 Métricas de Performance para Classificadores

Nesta seção, serão descritas algumas das principais métricas utilizadas para avaliar a performance de classificadores que serão aplicadas em seções posteriores.

2.3.1 Matriz de Confusão e Acurácia

Todas as métricas descritas aqui podem ser calculadas a partir da chamada Matriz de Confusão. Este artifício é utilizado para avaliar performance no contexto de aprendizagem

Tabela 1: Matriz de confusão para o caso binário

Classe Real	Classificado como Positivo	Classificado como Negativo
Positivo	VP	FP
Negativo	FN	VN

supervisionada. Consiste na representação de quatro situações da classificação: elementos que foram classificados corretamente como pertencendo uma determinada classe (Verdadeiros Positivos – VP), os que foram classificados corretamente como não pertencentes a classe (Verdadeiros Negativos – VN), aqueles que foram classificados incorretamente como pertencentes a classe (Falsos Positivos – FP) e finalmente, aqueles que foram classificados incorretamente como não pertencentes a classe (Falsos Negativos – FN) (SOKOLOVA; LAPALME, 2009). Esses quatro elementos constituem a matriz de confusão, que é mostrada na Tabela 1 para o caso binário na Tabela 1.

A **acurácia** é uma das medidas utilizadas para avaliar classificação. Esta consiste da razão entre as instâncias classificadas corretamente e o número total de instâncias:

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + FN + VN} \quad (2.3)$$

Apesar de ser uma das métricas mais utilizadas, a acurácia nem sempre é a métrica mais adequada, pois sofre de problemas, principalmente em conjuntos de dados com classes desbalanceadas. Por exemplo, considere um problema de 100 instâncias com duas classes, em que existem 95 instâncias da classe 1 e 5 instâncias da classe 2. Um classificador que simplesmente classifica todo elemento que recebe como pertencente a classe 1 teria uma acurácia de 95%. As medidas definidas nas próximas seções lidam melhor com esse tipo de situação.

2.3.2 Precisão e Recall

Precisão e **Recall** são duas métricas que medem, respectivamente, quantos exemplos classificados como positivos são realmente positivos e quantos exemplos positivos foram corretamente classificados como positivos (ZHOU, 2019):

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2.4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.5)$$

Essas medidas são normalmente utilizadas em conjunto, pois considera-se que uma é um complemento da outra, pois retratam duas situações diferentes.

2.3.3 Medida F_β

A medida F_β (RIJSBERGEN, 1979) é uma métrica que combina a Precisão em o Recall em uma único valor :

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{Recall} \cdot \text{Precisão}}{\beta^2 \cdot \text{Precisão} \cdot \text{Recall}} \quad (2.6)$$

Essa medida retorna valores altos quando ambos a Precisão e o Recall são altos. De forma geral, ela é bastante utilizada e é considerada mais informativa que a acurácia sobre a efetividade dos classificadores.

2.3.4 Kappa de Cohen

O Kappa de Cohen (COHEN, 1960) é uma métrica robusta para avaliar a performance de classificadores baseada na taxa de concordância entre dois observadores utilizando escalas nominais. No caso da classificação, o primeiro observador são as classes reais dos elementos, enquanto o segundo são as classes que foram preditas pelo classificador. Utilizando as células da matriz de confusão, o Kappa pode ser calculado da seguinte forma (CHICCO; WARRENS; JURMAN, 2021):

$$\kappa = \frac{2 \cdot (VP \cdot VN - FP \cdot FN)}{(VP + FP) \cdot (FP + VN) \cdot (VP + FN) \cdot (FN + VN)} \quad (2.7)$$

Os valores do Kappa, diferentemente das medidas anteriores, variam entre -1 (Classificação completamente errada) e 1(Classificação perfeita).

2.3.5 Métricas sobre Problemas Multi-Classe

Todas as métricas definidas anteriormente foram propostas para um caso de classificação com apenas duas classes. No entanto, claramente esse não é o caso em aplicações reais, que podem ter inúmeras classes. Existem duas estratégias para generalizar as métricas para o caso multi-classe (SOKOLOVA; LAPALME, 2009):

- Estratégia **macro-média**: Utilizando esta abordagem, as medidas são aplicadas

individualmente para cada classe c_1, \dots, c_n e então é feita uma média para obter o valor final para a métrica.

- Estratégia **micro-média**: Nesta abordagem, primeiramente é feito uma contagem para obter valores acumulados de VP, VN, FP e FN. E então, a métrica é aplicada sobre esses valores acumulados.

A grande diferença entre essas duas abordagens é que a abordagem de macro-média trata cada classe igualmente, enquanto a estratégia de micro-média favorece as classes maiores. Para futura referência, neste trabalho adotaremos a abordagem de **macro-média** sempre que aplicarmos alguma dessas métricas em problemas multi-classe.

2.4 Testes Estatísticos

Para validar os resultados obtidos em experimentos de diversas naturezas, normalmente são utilizados testes estatísticos. Nesta tese, também será adotada essa metodologia para validar os resultados obtidos com os algoritmos propostos.

Teste estatístico (ou testes de hipótese) é um procedimento estatístico que permite realizar inferências sobre uma determinada população através da rejeição de hipóteses realizadas sobre a mesma. O teste funciona criando duas hipóteses sobre a população: a **hipótese nula** (H_0) e a **hipótese alternativa** (H_1). Normalmente a conclusão que se quer chegar é a hipótese alternativa, que é um complemento da hipótese nula. O teste então provê meios para causar a rejeição da hipótese nula, levando à conclusão que a hipótese alternativa é verdadeira (DANIEL; CROSS, 2018).

Quando aplicado, o teste gera um valor denominado p -valor. Dizemos que podemos rejeitar a hipótese nula quando esse p -valor é inferior a um valor denominado nível de significância (α). Os valores mais comuns para α são 0.01, 0.05 e 0.1. Neste trabalho adotaremos o $\alpha = 0.05$, o que representa um nível de confiança de 95%.

Para serem utilizados, muitos testes estatísticos assumem que determinada população obedece a alguma distribuição estatística (como por exemplo, a distribuição normal). Testes que não realizam tal suposição são denominados testes **não-paramétricos**. Neste trabalho, utilizaremos dois testes não paramétricos para validar os experimentos: o teste de Friedman e o teste de Nemenyi.

O teste de Friedman é uma análise de variância utilizada para detectar se um número de procedimentos aplicados em uma determinada população tiveram efeitos diferentes. As

hipóteses (DANIEL; CROSS, 2018) deste teste são:

- H_0 : Os procedimentos produziram os mesmos efeitos.
- H_1 : Pelo menos um dos procedimentos produziu efeito diferente dos demais procedimentos.

Como teste post-hoc (ou seja, que será aplicado após o teste de Friedman) existem muitas alternativas como o teste de Nemenyi (NEMENYI, 1963) e o teste de Conover (CONOVER, 1999). Neste trabalho, será utilizado o teste de Nemenyi. Este teste é baseado no conceito de ranqueamento. No contexto de classificação, este teste inicialmente ranqueia os classificadores para cada conjunto de dados separadamente, com o melhor classificador recebendo o ranque 1, o segundo recebe o ranque 2, Caso ocorra um empate, é calculado o valor médio dos ranques.

O teste então calcula o ranque médio de cada classificador. A diferença de performance de dois classificadores é estatisticamente significativa se a média dos ranques correspondentes for maior que a diferença crítica (CD):

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (2.8)$$

onde k é o número de algoritmos, N é o número de conjuntos de dados e q_α é um valor baseado na estatística *Studentized range* (DEMŠAR, 2006).

Para tornar a visualização dos resultados do pareamento do teste de Nemenyi mais visual, pode-se utilizar o chamado Diagrama de Distância Crítica (Diagrama CD). Este diagrama consiste em um eixo em que são projetados os valores médios dos ranques dos classificadores. Se uma linha cruza dois ou mais métodos, significa que a diferença entre estes métodos não é estatisticamente significativa. Podemos visualizar um exemplo deste diagrama na Figura 5.

Neste exemplo, a diferença de desempenho entre os classificadores 2 e 4 não foi estatisticamente significativa. Além disso, a performance do classificador 1 foi a melhor e a do classificador 3, a pior.

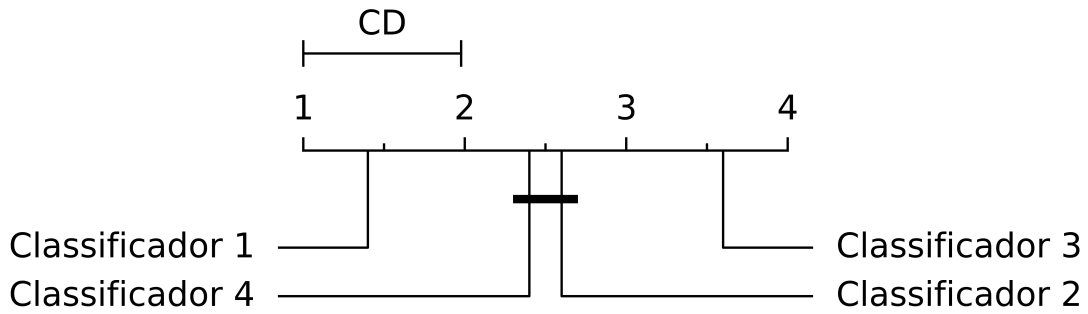


Figura 5: Exemplo de diagrama CD comparando a performance de 4 classificadores

2.5 Conceitos Matemáticos

Nesta seção, serão apresentados conceitos matemáticos presentes na literatura que serão utilizados ao longo deste trabalho.

Definição 2.1 Uma função F é dita ser uma **função de fusão** se for do tipo $F : [0, 1]^n \rightarrow [0, 1]$ para $n \in \mathbb{N}$, $n \geq 2$.

Definição 2.2 Sejam F_1 e F_2 duas funções de fusão. Dizemos que F_1 **domina** F_2 se:

$$\forall x_1, \dots, x_n \in [0, 1] : F_1(x_1, \dots, x_n) \geq F_2(x_1, \dots, x_n) \quad (2.9)$$

Definição 2.3 (BUSTINCE et al., 2018) Seja $\vec{r} = (r_1, \dots, r_n)$ um vetor n -dimensional, tal que $\vec{r} \neq \vec{0}$. Uma função de fusão $F : [0, 1]^n \rightarrow [0, 1]$ é \vec{r} -crescente se para todos os pontos $(x_1, \dots, x_n) \in [0, 1]^n$ e para todo $c > 0$ tal que $(x_1 + cr_1, \dots, x_n + cr_n) \in [0, 1]^n$ temos que :

$$F(x_1 + cr_1, \dots, x_n + cr_n) \geq F(x_1, \dots, x_n) \quad (2.10)$$

Definição 2.4 Seja $F : [0, 1]^n \rightarrow [0, 1]$ e \vec{r} um vetor real não nulo. F é ordenado direcionalmente (OD) \vec{r} -crescente, se para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$, e qualquer permutação σ de $N = \{1, \dots, n\}$ e $c > 0$ tal que $\vec{x}_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)}) \in L^n([0, 1])$, $\vec{x}_\sigma + c\vec{r} = (x_{\sigma(1)} + cr_1, \dots, x_{\sigma(n)} + cr_n) \in L^n([0, 1])$, onde $L^n([0, 1]) = \{(a_1, \dots, a_n) \in [0, 1]^n : a_1 \geq \dots \geq a_n\}$, verifica-se que $F(\vec{x}) \leq F(\vec{x} + c\vec{r}_{\sigma^{-1}})$.

Exemplo 2.1 Em (BUSTINCE et al., 2018, Exemplo 3.2) são apresentados vários exemplos de funções OD (\vec{r})-crescentes.

Por outro lado, considere $F(x_1, x_2, x_3) = x_k$ onde $k = \lceil x_1 + x_2 + x_3 \rceil$, com a convenção que $x_0 = 0$, $\vec{r} = (0, 0, 1)$, $\vec{x} = (0.7, 0.9, 0.4)$, $c = 0.1$ e a permutação seja: $\sigma(1) = 2$, $\sigma(2) = 1$ e $\sigma(3) = 3$. Então, $\vec{x}_\sigma = (0.9, 0.7, 0.4) \in L^n([0, 1])$, $\vec{x}_\sigma + c\vec{r} = (0.9 + 0.1 * 0, 0.7 + 0.1 * 0, 0.4 + 0.1 * 1) = (0.9, 0.7, 0.5) \in L^3([0, 1])$. Já que $F(\vec{x}) = x_2 = 0.9$ e $F(\vec{x} + 0.1\vec{r}_{\sigma^{-1}}) = F(0.7 + 0.1r_2, 0.9 + 0.1r_1, 0.4 + 0.1r_3) = F(0.7, 0.9, 0.5) = 0.5$, então F não é OD $(0, 0, 1)$ -crescente.

Definição 2.5 (SESMA-SARA et al., 2019) Seja $F : [0, 1]^n \rightarrow [0, 1]$ e \vec{r} um vetor real não nulo. F é fortemente ordenado direcionalmente (SOD) \vec{r} -crescente, se para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$, permutação σ de N e $c > 0$ tal que $\vec{x}_\sigma \in L^n([0, 1])$ e $\vec{x}_\sigma + c\vec{r} \in [0, 1]^n$ verifica-se que $F(\vec{x}) \leq F(\vec{x} + c\vec{r}_{\sigma^{-1}})$.

Claramente, ser SOD \vec{r} -crescente implica em ser OD \vec{r} -crescente, portanto a função F do Exemplo 2.1 também não é SOD \vec{r} -crescente para $\vec{r} = (0, 0, 1)$.

Definição 2.6 (LUCCA et al., 2016) Uma função de pré-agregação é uma função $F : [0, 1]^n \rightarrow [0, 1]$ que é \vec{r} -crescente para algum vetor real n -dimensional $\vec{r} = (r_1, \dots, r_n)$ e satisfaz as seguintes condições de borda: $F(0, \dots, 0) = 0$ and $F(1, \dots, 1) = 1$.

Definição 2.7 (BELIAKOV et al., 2007) Uma função de agregação é uma função $f : [0, 1]^n \rightarrow [0, 1]$ que mapeia o cubo unitário n -dimensional no intervalo unitário com as seguintes propriedades:

(A1) f é não decrescente¹

(A2) $f(0, \dots, 0) = 0$ e $f(1, \dots, 1) = 1$

Funções de agregação são funções de pré-agregação que são \vec{r} -crescentes para cada vetor positivo \vec{r} (BUSTINCE et al., 2016).

Definição 2.8 (BELIAKOV et al., 2007) Seja f uma função de agregação, dizemos que f é

1. Uma **função de média** se $\min \leq f \leq \max$;

2. **Conjuntiva** se $f \leq \min$

¹Uma função f é não decrescente se $\forall x, y \in \mathbb{R}, x < y \implies f(x) \leq f(y)$

3. **Idempotente** se $f(t, \dots, t) = t$ para todo $t \in [0, 1]$;
4. **Simétrica** se $f(x_1, \dots, x_n) = f(x_{1p}, \dots, x_{np})$, para qualquer permutação (x_{1p}, \dots, x_{np}) ;
5. **Estritamente monotônica** se $f(x_1, \dots, x_n) < f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n)$ para cada $(x_1, \dots, x_n) \in [0, 1]^n$ e $y \in [0, 1]$ tal que $x_i < y$;
6. Tem um **elemento aniquilador** se existe um $a \in [0, 1]$ tal que

$$f(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n) = a$$

para qualquer $i \in \{1, \dots, n\}$.

7. Tem um **elemento neutro** se existe $e \in [0, 1]$ tal que $f(e, \dots, e, x, e, \dots, e) = x$ para cada $x \in [0, 1]$.

Observação 2.1 *Uma função de agregação é uma função de média se, e somente se é idempotente (BELIAKOV et al., 2007).*

Cópulas são uma classe de operações definidas por A. Sklar (1959) que são extensamente utilizadas em probabilidade e estatística (ALSINA; SCHWEIZER; FRANK, 2006) e em outras áreas. Para o caso específico de cópulas de aridade 2, pode-se definir estas funções da seguinte forma:

Definição 2.9 (SKLAR, 1959; ALSINA; SCHWEIZER; FRANK, 2006) *Uma cópula é uma função $C : [0, 1]^2 \rightarrow [0, 1]$ com 0 como elemento aniquilador, 1 como elemento neutro e:*

$$C(x_1, y_1) - C(x_2, y_1) - C(x_1, y_2) + C(x_2, y_2) \geq 0 \quad (2.11)$$

sempre que $x_1 \leq x_2, y_1 \leq y_2$ e $x_1, x_2, y_1, y_2 \in [0, 1]$.

Observação 2.2 (GRABISCH et al., 2011) *Toda cópula é uma função de agregação conjuntiva.*

Definição 2.10 *Uma Cópula é positiva se $C(x, y) = 0$ implica que $x = 0$ ou $y = 0$. Iremos denotar o conjunto de todas as cópulas positivas por \mathcal{PC} .*

Funções de *overlap* são uma classe de funções de agregação introduzidas por Bustince et. al (2010a) para utilização em aplicações de visão computacional. Desde então, diferentes trabalhos (BEDREGAL et al., 2017; ELKANO et al., 2016; NOLASCO et al., 2019) tem expandido e utilizado estas funções em outras áreas, inclusive relacionando estas com outros tipos como as Cópulas.

Definição 2.11 (BUSTINCE et al., 2010a) Um **overlap** é uma função $O : [0, 1]^2 \rightarrow [0, 1]$ satisfazendo, para cada $x, y, z \in [0, 1]$, as seguintes propriedades:

- (O1) $O(x, y) = O(y, x)$ para cada $x, y \in [0, 1]$ (Comutatividade);
- (O2) $O(x, y) = 0 \iff xy = 0$;
- (O3) $O(x, y) = 1 \iff xy = 1$;
- (O4) $O(x, z) \leq O(y, z)$ para cada $x, y, z \in [0, 1]$ (Monotonicidade);
- (O5) O é contínuo.

Visto que a continuidade não é uma propriedade indispensável, em (PAIVA et al., 2021b) foi introduzida uma versão mais geral dos *overlaps*, chamada de *quasi-overlaps*, que têm as mesmas propriedades (O1)–(O4) dos *overlaps*, exceto a continuidade. De fato, existem situações que a continuidade não é necessária. Formalmente, pode-se prover a seguinte definição para os *quasi-overlaps*:

Definição 2.12 (MEZZOMO et al., 2020) Um **quasi-overlap** é uma função $O : [0, 1]^2 \rightarrow [0, 1]$ satisfazendo as propriedades (O1)–(O4) da Definição 2.11.

Iremos denotar por \mathcal{O} e \mathcal{QO} o conjunto de todos os *overlaps* e todos as funções de *quasi-overlaps*, respectivamente.

Na Tabela 2 estão definidos alguns exemplos de funções definidas anteriormente, com algumas dessas sendo usadas posteriormente neste trabalho.

2.5.1 Medidas Fuzzy

Uma medida *fuzzy* é um modelo matemático que atribui valor a cada possível combinação de valores de entrada. Apesar de alguns tipos serem computacionalmente custosos,

Tabela 2: Tabela com exemplos de diferentes tipos de cópulas, *overlaps* e *quasi-overlaps*

Nomenclatura	Definição	Classificação
O_σ	$O_\sigma(x, y) = \begin{cases} 0.1 + 0.8 * x^2 * y^2 & \text{se } x * y \notin \{0, 1\} \\ x * y & \text{caso contrário} \end{cases}$	<i>Quasi-Overlap</i> (e não é <i>Overlap</i> nem <i>Cópula</i>)
O_B	$O_B(x, y) = \min(x * \sqrt{y}, \sqrt{x} * y)$	<i>Overlap</i> e <i>Cópula</i>
O_C	$O_C(x, y) = \begin{cases} 0.1 + 0.8 * x * y & \text{se } x * y \notin \{0, 1\} \\ x * y & \text{caso contrário} \end{cases}$	<i>Quasi-Overlap</i> (e não é <i>Overlap</i> nem <i>Cópula</i>)
O_F	$O_F(x, y) = \frac{x * y + \min(x, y)}{2}$	<i>Overlap</i> e <i>Cópula</i>
O_2	$O_2(x, y) = \begin{cases} \frac{2 * x * y}{x + y} & \text{se } x + y \neq 0 \\ 0 & \text{caso contrário} \end{cases}$	<i>Overlap</i> (e não é <i>Cópula</i>)
O_T	$O_T(x, y) = x^2 y^2$	<i>Overlap</i> (e não é <i>Cópula</i>)

visto que em grande parte dos sistemas existem 2^n possíveis valores de entrada (BELIAKOV; JAMES; WU, 2020), simplificações deste modelo tem sido propostas com suposições que tornam o cálculo menos custoso.

Desde então, elas tem sido usadas para gerar duas classes importantes de funções de agregação do tipo média, as integrais de Choquet e de Sugeno, que incluem a classe das medias ponderadas ordenadas (OWA) (TORRA; NARUKAWA, 2007).

Definição 2.13 *Seja $N = \{1, \dots, n\}$ para um n arbitrário. Uma função $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ é dita ser uma **medida fuzzy** se as seguintes propriedades são satisfeitas:*

1. *Se $X \subseteq Y$, então $\mathbf{m}(X) \leq \mathbf{m}(Y)$ para todo $X, Y \subseteq N$*
2. $\mathbf{m}(\emptyset) = 0$
3. $\mathbf{m}(N) = 1$

A seguir, serão apresentados alguns exemplos de medidas *fuzzy* que estão disponíveis na literatura.

Exemplo 2.2 *Seja $N = \{1, \dots, n\}$ e $X \subseteq N$ e p, w dois vetores de pesos associados a X . Temos os seguintes exemplos de medidas fuzzy:*

- (BUSTINCE et al., 2016) A **medida fuzzy uniforme** é a função:

$$\mathbf{m}_U(X) = \frac{|X|}{n} \quad (2.12)$$

- A **medida fuzzy relativa** é a função:

$$\mathbf{m}_R(X) = \frac{\sum_{x \in X} x}{\sum_{i=1}^n i} \quad (2.13)$$

- *A medida fuzzy do produto é a função:*

$$\mathbf{m}_{\Pi}(X) = \frac{\prod_{x \in X} x}{\prod_{i=1}^n i} \quad (2.14)$$

- (BUSTINCE et al., 2016) *A medida fuzzy potência é a função:*

$$\mathbf{m}_p(X) = \left(\frac{|X|}{n}\right)^q, \text{ com } q > 0 \quad (2.15)$$

- (LLAMAZARES, 2018). *A medida fuzzy ponderada é a função:*

$$\mathbf{m}_w(X) = \sum_{i \in X} p_i \quad (2.16)$$

- (LLAMAZARES, 2018) *A medida fuzzy OWA é a função:*

$$\mathbf{m}_{owa}(X) = \sum_{i=1}^{|X|} w_i \quad (2.17)$$

Exemplo 2.3 (LI; WANG; CHEN, 2015) *Um outro exemplo de classe de medida fuzzy é a classe de medidas fuzzy λ . Sejam $A, B \in \mathcal{P}(X)$ tal que $A \cap B = \emptyset$, \mathbf{m} é dita ser uma medida fuzzy λ se satisfaz:*

$$\mathbf{m}(A \cup B) = \mathbf{m}(A) + \mathbf{m}(B) + \lambda \mathbf{m}(A)\mathbf{m}(B) \quad (2.18)$$

$\forall A, B \in \mathcal{P}(X)$ com $\lambda \geq -1$

Os valores para cada $\mathbf{m}(A)$, onde $A \subseteq \mathcal{P}(X)$, podem ser calculados pela seguinte fórmula:

$$\mathbf{m}\left(\bigcup_{x \in A} \{x\}\right) = \frac{1}{\lambda} \left(\prod_{x \in A} (1 + \lambda \mathbf{m}(\{x\})) - 1 \right), \lambda \neq 0 \quad (2.19)$$

Além disso, o coeficiente λ pode ser encontrado ao resolver o seguinte polinômio

$$\lambda + 1 = \prod_{x \in A} (1 + \lambda \mathbf{m}(\{x\})) \quad (2.20)$$

que pode ser resolvido em $(-1, 0)$ ou $(0, \infty)$.

Definição 2.14 (BELIAKOV; JAMES; WU, 2020) *Sejam $X, Y \subseteq N = \{1, \dots, n\}$ conjuntos*

finitos. Uma medida fuzzy \mathbf{m} é dita ser **simétrica** se:

$$|X| = |Y| \implies \mathbf{m}(X) = \mathbf{m}(Y)$$

Em outras palavras, o valor de \mathbf{m} depende apenas da cardinalidade do conjunto a qual a medida fuzzy foi aplicada.

Observação 2.3 Das medidas fuzzy apresentadas no Exemplo 2.2, apenas as medidas fuzzy uniforme e potência são simétricas.

Definição 2.15 Seja $X \subseteq N = \{1, \dots, n\}$ um conjunto finito. Uma medida fuzzy \mathbf{m} é dita ser **estritamente limitada** se:

$$\mathbf{m}(X) = 0 \implies X = \emptyset \text{ e } \mathbf{m}(X) = 1 \implies X = N$$

2.5.2 Integral de Choquet

A Integral de Choquet, proposta por Gustave Choquet (1954) é uma generalização da Integral de Lebesgue que compartilha muita das propriedades com esta. Estas integrais tem apresentado excelentes resultados em diferentes áreas (DIMURO et al., 2020a; BUSTINCE et al., 2021) como avaliação (SHIEH; WU; LIU, 2009), reconhecimento facial (KARCZMAREK; KIERSZTYN; PEDRYCZ, 2018) e previsão do tempo (ABDEL-NASSER; MAHMOUD; LEHTONEN, 2020).

A Integral de Choquet é definida com respeito a medidas fuzzy. Neste trabalho usaremos a versão discreta da Integral de Choquet e de suas variantes, visto que os conjuntos de dados utilizados serão finitos e discretos. Porém, a versão original da integral de Choquet pode ser encontrada facilmente na literatura.

O principal propósito das integrais de Choquet é combinar as entradas de uma forma que leve em consideração a disposição dos elementos da entrada, ou seja, a combinação de elementos fornecida, além dos fatores que normalmente são utilizados em funções de agregação como a importância de cada indivíduo específico (no caso da média ponderada), sua magnitude (como é o caso da OWA) (BELIAKOV et al., 2007).

Definição 2.16 (LUCCA et al., 2017) Seja $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy. A integral de Choquet discreta é a função $\mathcal{C}_{\mathbf{m}} : [0, 1]^n \rightarrow [0, 1]$, definida para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$, por:

$$\mathcal{C}_m(\vec{x}) = \sum_{i=1}^n \mathbf{m}(A_{(i)}) * (x_{(i)} - x_{(i-1)}) \quad (2.21)$$

onde $(x_{(1)}, \dots, x_{(n)})$ é uma permutação crescente de \vec{x} , por convenção, $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$. Aplicando a distributividade, também podemos obter a integral de Choquet expandida, que será a forma utilizada nesta tese:

$$\mathcal{C}_m(\vec{x}) = \sum_{i=1}^n (x_{(i)} * \mathbf{m}(A_{(i)}) - x_{(i-1)} * \mathbf{m}(A_{(i)})) \quad (2.22)$$

Lucca et. al (2017) introduziu as CC-Integrais, que são funções de agregação baseadas na Integral de Choquet que utilizam Cópulas para substituir o produto que existe na Equação (2.21), obtendo bons resultados em comparação a Integral de Choquet discreta clássica. Ela será definida a seguir.

Definição 2.17 (LUCCA et al., 2017) *Seja $N = \{1, \dots, n\}$ um conjunto finito, $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy e $C : [0, 1]^2 \rightarrow [0, 1]$ uma cópula. A Integral de Choquet baseada em cópula (CC-Integral) é definida como a função $\mathcal{C}_m^C : [0, 1]^n \rightarrow [0, 1]$, dada para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$ por:*

$$\mathcal{C}_m^C(\vec{x}) = \sum_{i=1}^n (C(x_{(i)}, \mathbf{m}(A_{(i)})) - C(x_{(i-1)}, \mathbf{m}(A_{(i)}))) \quad (2.23)$$

onde $(x_{(1)}, \dots, x_{(n)})$ é uma permutação crescente de \vec{x} , por convenção, $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$.

Esse conceito foi expandido em (LUCCA et al., 2019), onde foi proposta a Integral $C_{F_1 F_2}$. Nestas integrais, as Cópulas presentes na Equação (2.23) são substituídas por funções de fusão F_1 e F_2 , que podem ser distintas. Essa variante também possui algumas restrições que serão descritas a seguir.

Definição 2.18 (LUCCA et al., 2019) *Seja $N = \{1, \dots, n\}$ um conjunto finito, $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy simétrica e $F_1, F_2 : [0, 1]^2 \rightarrow [0, 1]$ duas funções de fusão tal que F_1 domina sobre F_2 . A Integral de Choquet $F_1 F_2$ (Integral $C_{F_1 F_2}$) é definida como a função $\mathcal{C}_m^{(F_1 F_2)} : [0, 1]^n \rightarrow [0, 1]$, dada para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$ por:*

$$\mathcal{C}_m^{(F_1 F_2)}(\vec{x}) = \min \left\{ 1, x_1 + \sum_{i=2}^n (F_1(x_{(i)}, \mathbf{m}(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}(A_{(i)}))) \right\} \quad (2.24)$$

onde $(x_{(1)}, \dots, x_{(n)})$ é uma permutação crescente de \vec{x} , por convenção, $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$.

No entanto, diferentemente das CC-integrais, as integrais $C_{F_1 F_2}$ nem sempre são funções de agregação, de fato algumas delas nem sequer são pré-agregações (DIMURO et al., 2020a).

2.6 Considerações Finais

Neste capítulo, foram apresentados conceitos fundamentais que serão utilizados no decorrer desta tese. Inicialmente, foi apresentado conceitos básicos de aprendizagem de máquina e alguns dos seus principais métodos, que serão posteriormente utilizados para construir comitês de classificadores, que foram explicados em seguida. Finalmente, foi apresentado o arcabouço matemático que será necessário para construção das contribuições desta tese.

3 Trabalhos Correlatos

Neste capítulo, apresentaremos trabalhos que têm objetivos ou metodologias similares ao desta tese, de forma a examinar o estado da arte, através de uma revisão narrativa. Inicialmente, serão explorados outros trabalhos que propuseram generalizações da Integral de Choquet. Em seguida, serão mostrados trabalhos que adotaram o modelo de comitês de classificadores, com foco naqueles que propuseram novos métodos de combinação e principalmente, aqueles que utilizaram Integrais de Choquet ou alguma de suas generalizações.

Primeiramente, é importante deixar claro qual a metodologia será utilizada para seleção dos trabalhos aqui apresentados. Visto que não é o objetivo desta tese fazer uma profunda revisão no estado da arte de técnicas de combinação em comitês de classificadores, onde neste caso seria adequado uma revisão sistemática, utilizaremos uma revisão narrativa.

Revisões narrativas são sínteses de informações previamente publicadas em um formato condensado que sumariza o conteúdo de cada trabalho (GREEN; JOHNSON; ADAMS, 2006). Ao contrário das revisões sistemáticas, não exige um protocolo rígido para sua confecção e a busca das fontes não é pré-determinada e específica, com a seleção dos artigos fazendo parte da percepção do autor relativa ao tema (CORDEIRO et al., 2007).

3.1 Generalizações da Integral de Choquet

Neste trabalho, já foram discutidas e apresentadas duas generalizações da integral de Choquet. A CC-Integral, proposta por Lucca et al. (2017), substituiu o produto presente nas Integrais por Copulas. Ainda neste trabalho, os autores testaram esta integral no Contexto de classificadores baseados em regras *fuzzy* e obtiveram resultados que mostraram que foi uma abordagem competitiva quando comparada com outras técnicas semelhantes.

O segundo método que já foi apresentado nesta tese foi a integral $C_{F_1 F_2}$ (LUCCA et al.,

2019). Esta generalização da integral de Choquet discreta expandida substituiu o produto entre a medida *fuzzy* e os elementos por duas funções de fusão que podem ser distintas, mas com uma condição de dominância de uma função sobre a outra. Esta abordagem utiliza uma única medida *fuzzy* simétrica. Os autores ainda testaram este modelo novamente no contexto de classificadores baseados em regras *fuzzy* e obtiveram resultados competitivos contra outros métodos da área como o algoritmo FURIA (HÜHN; HÜLLERMEIER, 2009).

Dimuro et al. (2020b) propôs uma generalização da integral $C_{F_1 F_2}$ denominada $gC_{F_1 F_2}$. Nesta abordagem, o produto da integral de Choquet discreta expandida foi substituído por um par de funções de pseudo pré-agregação. Neste trabalho, os autores se preocuparam em estabelecer as condições necessárias para que a integral esteja bem definida no intervalo $[0, 1]$, além de outras condições. Além disso, aplicaram a redução de dimensionalidade para lidar com o problema de valores repetidos no vetor de entrada. É importante destacar que a medida *fuzzy* utilizada nesta generalização não necessariamente é simétrica.

Em (BUSTINCE et al., 2021), foram propostas as chamadas d-Choquet integrals, uma generalização das Integrais de Choquet em que a subtração entre os elementos da integral foi substituída por uma medida de dissimilaridade restrita. Posteriormente, em (TAKÁČ et al., 2021) os autores generalizaram a noção de d-Choquet Integrals, propondo uma nova generalização no contexto de valores intervalares, chamada dg-Choquet integral. Esse modelo também foi aplicado em um contexto de combinação de classificadores. Porém neste caso, como uma camada para combinar o resultados obtidos por dois comitês, e não como método de combinação dentro do comitê.

3.2 Comitês de Classificadores

Nesta seção, serão mostrados alguns trabalhos que tratam especificamente com o modelo de comitês de classificadores, focando principalmente na etapa de combinação. Visto que neste trabalhos não iremos utilizar a abordagem de seleção na camada de combinação, não serão explorados trabalhos que utilizam esta metodologia, portanto serão apresentados aqueles que utilizaram a abordagem de fusão. Inicialmente, serão mostrados alguns métodos de construções de comitês.

3.2.1 Métodos de Criação de Comitês de Classificadores

Uma das principais técnicas utilizadas para comitês de classificadores é o chamado *boosting*, basicamente esta técnica consiste em agrupar classificadores fracos em um clas-

sificador forte. Um classificador é dito ser **fraco** quando ele tem baixa performance, já um classificador **forte** tem uma boa acurácia. Um dos trabalhos que explorou este tipo de técnica foi o Kumar e Roy (2016), que usou um algoritmo chamado *LogitBoost* com árvores de Decisão para criar um comitê homogêneo para previsão de falência em empresas, obtendo bons resultados.

Em (BARZEGAR et al., 2018), os autores também utilizaram um algoritmo de boosting, com a aplicação de controle de qualidade da água. Os autores escolheram o algoritmo *LS-Boosting* em comitês com versões diferentes de redes neurais, onde os autores concluíram que a técnica utilizando comitês teve um desempenho superior aos classificadores testados individualmente. Já em (KADKHODAEI; MOGHADAM; DEGHAN, 2020), os autores propuseram um algoritmo chamado HBoost baseado em boosting, que cria comitês heterogêneos de 10 classificadores. Em geral o algoritmo proposto teve resultados superiores a outras técnicas propostas, como o AdaBoost, Stacking e Random Forest.

Uma das abordagens mais utilizadas recentemente para construção de comitês é o XGBoost (CHEN; GUESTRIN, 2016). Esta técnica é uma extensão dos algoritmos de Boosting baseado em árvores. Esta técnica se tornou popular por consumir menos recursos que abordagem tradicional, além de apresentar resultados superiores.

Uma outra conhecida técnica de classificação é o *bagging*. Esta técnica utiliza uma amostragem com reposição para criar classificadores treinados com partes diferentes do conjunto de dados. Um trabalho que usou esta técnica foi (LUO, 2018), onde um algoritmo de Bagging foi utilizado para criar comitês homogêneos de Redes Neurais, *Support Vector Machines* – SVM, Árvores de Decisão e k-NN. De acordo com (ZHOU, 2019), uma das técnicas de combinação mais utilizadas em ensembles que usam *bagging* é a votação por maioria.

Um outro método proposto recentemente é o META-DES (CRUZ et al., 2015). Este é baseado na abordagem de seleção para comitês. O META-DES funciona através da atribuição de um critério para julgar se um classificador é competente o suficiente para fazer parte do comitê. Essa atribuição é feita a partir de um meta-classificador, que estima a competência do classificador em relação ao conjunto de dados e julga se determinado classificador irá fazer parte do comitê.

3.2.2 Métodos de Combinação baseados em Fusão

Podemos dividir os métodos de combinação baseados em fusão em duas principais estratégias: a combinação *hard* e *soft* (MOHANDÉS; DERICHE; ALIYU, 2018). Na combinação *hard*, apenas o valor da classe obtida de cada classificador é levada em consideração no momento da combinação. Na combinação *soft*, a combinação é realizada sobre um vetor obtido de cada classificador, que contém pontuações (como por exemplo, um vetor de probabilidade) que representam a confiança do classificador com relação ao pertencimento de um determinado elemento a cada classe.

Como o principal foco desta tese é propor um método de combinação baseado em funções de agregação, nesta seção serão mostrados alguns trabalhos presentes na literatura que possuem a mesma abordagem. Tipicamente estes métodos caem na categoria de combinação *soft*. Nestes sistemas, em geral, a soma ponderada da saída de todos os classificadores base é calculada, levando em conta pesos predefinidos anteriormente, ou calculados em tempo de execução (ALZUBI et al., 2018).

3.2.2.1 Combinação de fusão *hard*

A votação por maioria é uma técnica simples mas constantemente utilizada em trabalhos que utilizam comitês de classificadores. Esta técnica tem a grande vantagem de acarretar em um custo computacional praticamente insignificante e apresentar bons resultados. Por exemplo, o trabalho (ATALLAH; AL-MOUSA, 2019) criou um comitê heterogêneo de classificadores utilizando o método de combinação de votação por maioria consistindo de Gradiente Descendente, k-NN, Random Forest e Regressão Logística e obteve taxas de acerto superiores a 90% para detecção de doenças do coração em pacientes.

Um outro tipo de votação presente em comitês é a votação por pluralidade. Diferentemente da votação por maioria, que requer que o vencedor receba pelo menos 50% dos votos, a votação por pluralidade seleciona o candidato que recebeu mais votos como o vencedor. Em (LIN et al., 2003), os autores demonstraram que a votação por pluralidade tem desempenho superior a votação por maioria. Outros trabalhos expandiram a ideia de votação, como por exemplo o trabalho de (DOGAN; BIRANT, 2019), propôs um sistema de votação baseado em pesos que obteve excelentes resultados quando comparado com outras técnicas.

3.2.2.2 Combinação de fusão *soft*

Em seu trabalho, (COSTA et al., 2018) utilizou funções *General Mixture* – GM, para combinar a saída dos classificadores em um comitê heterogêneo. Este sistema se baseia na geração de pesos dinâmicos definidos em tempo de execução para cada classificador, que são calculados a partir das funções GM. Em seguida esses pesos são usados para calcular a decisão final, utilizando um sistema similar a uma média ponderada, utilizando os pesos calculados previamente pelo modelo, onde classificadores que tiverem uma melhor performance terão um peso maior na decisão. Os autores compararam seu método com outros métodos estabelecidos na literatura como o RandomForest e o XGBoost e obtiveram resultados superiores nas bases selecionadas.

Em (ALZUBI et al., 2018), propôs uma nova abordagem para geração de pesos para serem utilizados como combinação linear. Neste modelo, os pesos dos classificadores são ajustados repetidamente pelos próprios classificadores baseado em cálculos de incerteza, até que cheguem a um consenso sobre qual a configuração final de pesos. Esse sistema também usa um cálculo baseado em média ponderada.

Outras abordagens que podem ser encontradas na literatura incluem, por exemplo, uso de diversos tipos de Integrais *Fuzzy*, como a integral de Sugeno e a de Shilkret em conjunto com densidades *fuzzy* para realizar a combinação em um comitê de classificadores (PIZZI; PEDRYCZ, 2010). Além disso, este trabalho mostra que o desempenho do comitê foi superior ao dos classificadores individuais.

Em (LIU et al., 2017), é realizada uma avaliação da confiabilidade de cada classificador. A idéia é que classificadores menos confiáveis terão um peso menor na decisão. Os autores então utilizam a regra de Dempster-Shafer realizar a combinação dos resultados após os cálculos da confiabilidade e mostram que os resultados foram superiores a alguns outros métodos de fusão. Existem ainda versões de votação por maioria que se encaixam na categoria de combinação *soft*, como é o caso da votação por maioria *fuzzy*, proposta por Sá et al. (2020).

3.2.2.3 Métodos de Combinação com Integrais de Choquet

Nesta seção, serão mostrados especificamente trabalhos que utilizam Integrais de Choquet para realizar a combinação da saída dos classificadores. Note que essa abordagem também pode ser considerada uma combinação *soft*. Por exemplo, em (DEY et al., 2021), foi utilizada uma versão *fuzzy* da integral de Choquet para agregar a saída de um comitê

de classificadores para detecção de COVID-19 em imagens de raio-x de pacientes. Neste trabalho, os autores também propuseram uma abordagem para calcular os valores a serem utilizados pela medida *fuzzy* em tempo real.

O trabalho de (LI; WANG; CHEN, 2015) utiliza Integrais de Choquet para combinação em comitês homogêneos de SVMs utilizando também a técnica de *Bagging*. Este sistema funciona dividindo o conjunto de treino e oferecendo esse conjunto para os classificadores individualmente. Após a classificação, a saída destes classificadores é combinada utilizando uma integral de Choquet utilizando uma medida fuzzy λ .

Em (PACHECO; KROHLING, 2018), os autores construíram um comitê de classificadores utilizando redes neurais, e aplicaram a integral de Choquet para combinar as classes providas por cada classificador no comitê. Nesta abordagem, os autores utilizaram uma medida fuzzy baseada na entropia de *Shannon*. Os autores então aplicaram essa abordagem em bases de dados disponíveis *on-line* e obtiveram bons resultados.

Finalmente, em (CAO, 2012), os autores propõem o uso de um comitê de classificadores heterogêneo para previsão de dificuldades financeiras. Da mesma forma que os trabalhos anteriormente citados nesta seção, eles utilizam a integral de Choquet para combinar a opinião dos classificadores. Neste caso, os autores propõem uma medida *fuzzy* dinâmica baseada em uma densidade *fuzzy*, realizando ajustes em tempo de execução para ajustar o valor da importância dos classificadores baseada na confiança da classificação.

3.3 Considerações Finais

Neste capítulo, apresentamos diversos trabalhos presentes na literatura em que o tema proposto tem relação com o abordado nesta tese. Inicialmente, foi discutida técnicas de criação de comitês de classificadores, e em seguida foram descritas diferentes abordagens presentes na literatura de combinação das saídas dos classificadores em um comitê. Finalmente, foram mostrados trabalhos que utilizaram a Integral de Choquet. O que pode-se perceber destes trabalhos é que o seu grande foco é a proposição de uma medida *fuzzy* modificada ou até nova para a Integral de Choquet.

Neste trabalho, será proposta uma modificação na própria estrutura da Integral de Choquet, diferentemente dos trabalhos anteriormente citados, além de outras contribuições, que serão expostas em seções posteriores. Além disso, usaremos alguns dos métodos apresentados neste capítulo como comparação com o método proposto.

4 Contribuições Teóricas

Neste capítulo iremos apresentar as contribuições teóricas desta tese. Inicialmente, serão propostas duas novas classes de funções de agregação, chamadas *pseudo overlaps* e *pseudo-quasi overlaps*. Utilizando estas funções previamente citadas, serão definidas duas generalizações da integral de Choquet, a integral CQO e a Pseudo-Choquet Integral. Finalmente, será mostrada a Integral de Choquet Absoluta, que é a contribuição principal desta tese.

4.1 *Pseudo Overlaps*

Os *Pseudo Overlaps* são uma classe de funções de agregação que são funções de *overlaps* em que não é exigida a propriedade de comutatividade. A utilização de funções não comutativas é explorada em diferentes trabalhos e são extensamente utilizadas, por exemplo, na definição de operadores *fuzzy* não comutativos (DAI; CHENG, 2019).

Definição 4.1 *Uma função bivariada $PO : [0, 1]^2 \rightarrow [0, 1]$ é dita ser uma função pseudo overlap se, para cada $x, y, z \in [0, 1]$, satisfaz as seguintes condições:*

(PO1) $PO(x, y) = 0$ se, e somente se $xy = 0$

(PO2) $PO(x, y) = 1$ se, e somente se $xy = 1$

(PO3) $PO(x, z) \leq PO(y, z)$ para cada $x, y, z \in [0, 1]$

(PO4) PO é contínuo.

Iremos denotar por \mathcal{PO} o conjunto de todos os *pseudo overlaps*.

Exemplo 4.1 *As seguintes funções são exemplos de pseudo overlaps:*

1. $PO_{DB}^{k,\ell}(x, y) = \begin{cases} \frac{2x^k y^\ell}{x+y}, & \text{se } x + y \neq 0 \\ 0, & \text{se } x + y = 0 \end{cases}$, com inteiros $k, \ell \geq 1$;
2. $PO^{r,s}(x, y) = x^r y^s$, com $r, s > 0$ e $r, s \neq 1$;
3. $PO_{mM}^{p,q}(x, y) = \min(x, y) \max(x^p, y^q)$, com inteiros $p, q \geq 2$;
4. $PO(x, y) = \begin{cases} \frac{1+(2x-1)^2(2y-1)^4}{2}, & \text{se } x, y \in [0.5, 1] \\ \min(x, y), & \text{caso contrário.} \end{cases}$

Obviamente, todo *overlap* é um *pseudo overlap*. Um *pseudo overlap* é dito ser *próprio* se não é comutativo. Os seguintes teoremas mostram como transformar *pseudo overlaps* próprios em funções de *overlap*.

Teorema 4.1 *Seja $PO : [0, 1]^2 \rightarrow [0, 1]$ um pseudo overlap. Os mapeamentos $O_1, O_2 : [0, 1]^2 \rightarrow [0, 1]$ dados por:*

- $O_1(x, y) = PO(\min\{x, y\}, \max\{x, y\})$ e
- $O_2(x, y) = PO(\max\{x, y\}, \min\{x, y\})$

são duas funções de *overlap*.

Prova: De fato, se PO é um *pseudo overlap* e O_1 é dado por $O_1(x, y) = PO(\min\{x, y\}, \max\{x, y\})$, então temos que:

$$(O1): O_1(x, y) = PO(\min\{x, y\}, \max\{x, y\}) = PO(\min\{y, x\}, \max\{y, x\}) = O_1(y, x);$$

$$(O2): O_1(x, y) = 0 \text{ se, e somente se } PO(\min\{x, y\}, \max\{x, y\}) = 0 \text{ se, e somente se } \min\{x, y\} \max\{x, y\} = 0 \text{ se, e somente se } xy = 0;$$

(O3): análogo ao anterior;

(O4): para cada $x, y, z \in [0, 1]$ tal que $y \leq z$ temos que $\min\{x, y\} \leq \min\{x, z\}$ e $\max\{x, y\} \leq \max\{x, z\}$. Então, by (PO3), segue que

$$\begin{aligned} O_1(x, y) = PO(\min\{x, y\}, \max\{x, y\}) &\leq PO(\min\{x, z\}, \max\{x, z\}) \\ &= O_1(x, z). \end{aligned}$$

Portanto, PO é crescente;

(O5): direto de (PO4).

Portanto, O_1 é uma função de *overlap*. Similarmente, conclui-se que O_2 também é um *overlap*. \square

Teorema 4.2 *Seja $A : [0, 1]^2 \rightarrow [0, 1]$ uma função de agregação e PO um pseudo overlap. O mapeamento $O : [0, 1]^2 \rightarrow [0, 1]$ dado por $O(x, y) = A(PO(x, y), PO(y, x))$ é um overlap se, e somente se satisfaz:*

- (i) A é comutativo e contínuo
- (ii) $A(x, y) > 0$ sempre que $x, y > 0$
- (iii) $A(x, y) < 1$ sempre que $x, y < 1$.

Prova: Assumindo que O é um *overlap*, dado a maneira que foi definido, a comutatividade e continuidade de A seguem diretamente de (O1) e (O5). Além disso, por (O2) segue que $A(x, y) > 0$ sempre que $x, y > 0$. E por (O3) segue que $A(x, y) < 1$ sempre que $x, y < 1$. Alternativamente, suponha que A é uma função de agregação que satisfaz as propriedades (i), (ii) and (iii). Então, por (ii), a função O dada por $O(x, y) = A(PO(x, y), PO(y, x))$ satisfaz (O1) e (O5). Se $O(x, y) = A(PO(x, y), PO(y, x)) = 0$, então por (ii) $PO(x, y) \cdot PO(y, x) = 0$, que implica por (PO1) que $xy = 0$. Por outro lado, se $xy = 0$ então $PO(x, y) = PO(y, x) = 0$ e portanto, por (A2), $0 = A(PO(x, y), PO(y, x)) = O(x, y)$. Desta forma, O satisfaz (O2). De maneira similar, usando (iii) pode ser mostrado que O satisfaz (O3). Finalmente, por (PO3) e (A1), para cada $y, z \in [0, 1]$ tal que $y \leq z$, temos que

$$\begin{aligned} O(x, y) &= A(PO(x, y), PO(y, x)) \\ &\leq A(PO(x, z), PO(z, x)) \\ &= O(x, z). \end{aligned}$$

Portanto, O satisfaz (O4). \square

O seguinte resultado dá uma caracterização alternativa as funções de *pseudo overlap*

Lema 4.1 *A função bivariada $PO : [0, 1]^2 \rightarrow [0, 1]$ é um pseudo overlap se, e somente se*

$$PO(x, y) = \frac{f(x, y)}{f(x, y) + h(x, y)}$$

para algum $f, h : [0, 1]^2 \rightarrow [0, 1]$ tal que

- (i) f é crescente e h é decrescente;
- (ii) $f(x, y) = 0$ se, e somente se $xy = 0$;
- (iii) $h(x, y) = 0$ se, e somente se $xy = 1$;
- (iv) f e h são funções contínuas.

Prova: Análogo a (BUSTINCE et al., 2010b, Teorema 5). □

4.2 *Pseudo-Quasi Overlaps*

Uma extensão natural dos *quasi overlaps* e *pseudo overlaps* pode ser obtida ao definir uma classe de *Overlaps* não contínuos e não simétricos. Conforme discutido na Seção 2.5, a continuidade não é uma propriedade indispensável (SU; WANG, 2013). O relaxamento da propriedade de simetria já foi feito com outras classes de funções como as t-normas, t-conormas e uninormas, através da definição de pseudo t-normas, pseudo t-conormas e pseudo uninormas, que são versões não comutativas das funções originais (FLONDOR; GEORGESCU; IORGULESCU, 2001; SILVA et al., 2015)

Desta forma, nesta seção será proposta uma nova classe de funções de agregação resultado de um maior relaxamento da condição dos *overlaps*, removendo a necessidade de continuidade e simetria da definição. Chamaremos essa nova função de *pseudo-quasi overlap*:

Definição 4.2 *Um pseudo-quasi overlap é uma função $O_{PQ} : [0, 1]^2 \rightarrow [0, 1]$ satisfazendo as seguintes propriedades:*

$$(PQO1) \quad O_{PQ}(x, y) = 0 \iff xy = 0$$

$$(PQO2) \quad O_{PQ}(x, y) = 1 \iff xy = 1$$

$$(PQO3) \quad O_{PQ}(x, z) \leq O_{PQ}(y, z) \text{ para cada } x, y, z \in [0, 1] \text{ (Monotonicidade)}$$

Além disso,

1. O_{PQ} é um \wedge_t -morfismo se $O_{PQ}(x \wedge y, z) = O_{PQ}(x, z) \wedge O_{PQ}(y, z)$, para todo $x, y, z \in [0, 1]$;

2. O_{PQ} é um \wedge_r -morfismo se $O_{PQ}(x, y \wedge z) = O_{PQ}(x, y) \wedge O_{PQ}(x, z)$ para todo $x, y, z \in [0, 1]$;
3. O_{PQ} é um \wedge -morfismo se é um \wedge_l -morfismo e um \wedge_r -morfismo;
4. O_{PQ} é um \vee_l -morfismo se $O_{PQ}(x \vee y, z) = O_{PQ}(x, z) \vee O_{PQ}(y, z)$ para todo $x, y, z \in [0, 1]$;
5. O_{PQ} é um \vee_r -morfismo se $O_{PQ}(x, y \vee z) = O_{PQ}(x, y) \vee O_{PQ}(x, z)$ para todo $x, y, z \in [0, 1]$;
6. O_{PQ} é um \vee -morfismo se é um \vee_l -morfismo e \vee_r -morfismo;
7. 1 é um elemento l -neutro de O_{PQ} se $O_{PQ}(1, x) = x$ para todo $x \in [0, 1]$;
8. 1 é um elemento r -neutro de O_{PQ} se $O_{PQ}(x, 1) = x$ para todo $x \in [0, 1]$;
9. 1 é um elemento neutro de O_{PQ} se é um elemento l -neutro e r -neutro;
10. O_{PQ} é l -subneutro se $O_{PQ}(1, x) \leq x$ para todo $x \in [0, 1]$;
11. O_{PQ} é r -subneutro se $O_{PQ}(x, 1) \leq x$ para todo $x \in [0, 1]$;
12. O_{PQ} é subneutro se é l -subneutro e r -subneutro;
13. O_{PQ} é idempotente if $O_{PQ}(x, x) = x$ para todo $x \in [0, 1]$;
14. O_{PQ} é sub-idempotente if $O_{PQ}(x, x) \leq x$ para todo $x \in [0, 1]$.

Iremos denotar por \mathcal{PQO} o conjunto de todos os *pseudo-quasi overlaps*

Observação 4.1 Obviamente, $\mathcal{O} \subseteq \mathcal{QO} \subseteq \mathcal{PQO}$ e $\mathcal{O} \subseteq \mathcal{PO} \subseteq \mathcal{PQO}$. Tanto a l -subneutralidade como a r -subneutralidade implica na sub-idempotência.

Observação 4.2 Se um pseudo-quasi overlap tem elemento l -neutro ou r -neutro, ele necessariamente é 1 .

Proposição 4.1 Seja O_{PQ} um pseudo-quasi overlap. O_{PQ} é subneutro se, e somente se, O_{PQ} é conjuntivo.

Prova: (\Rightarrow) Claramente, $O_{PQ}(x, y) \leq O_{PQ}(x, 1) \wedge O_{PQ}(1, y) \leq x \wedge y$, para todo $x, y \in [0, 1]$.

(\Leftarrow) $O_{PQ}(x, 1) \leq x \wedge 1 = x$. □

Exemplo 4.2 *Toda função da forma*

$$f(x, y) = x^m y^n$$

com $m, n > 0$ é um pseudo-quasi overlap. Além disso, se $m \neq n$, f não é um quasi-overlap e portanto não é um overlap. Note que f é um \wedge -morfismo, \vee -morfismo e l -subneutro, mas não é nem r -subneutro nem idempotente.

Exemplo 4.3 *A função*

$$g(x, y) = \begin{cases} x & \text{se } x \leq y \\ \frac{x+y}{2} & \text{se } x \geq y > 0 \\ 0 & \text{se } y = 0 \end{cases}$$

é um pseudo-quasi overlap que não é contínuo, não é idempotente, e também não é um quasi-overlap. Além disso, 1 é um elemento r -neutro de g mas não é um elemento l -neutro, e portanto não é conjuntivo. g é um \wedge -morfismo e \vee -morfismo.

Um outro exemplo de uma família de pseudo-quasi overlaps idempotentes pode ser definido para cada $\lambda \in [0, 1]$ a partir da função:

$$h_\lambda(x, y) = \begin{cases} 0 & \text{se } xy = 0 \\ \lambda x + (1 - \lambda)y & \text{caso contrário} \end{cases}$$

Observe que dado um pseudo-quasi overlap O_{PQ} , seu simétrico, ou seja $\overleftrightarrow{O}_{PQ}(x, y) = O_{PQ}(y, x)$, também é um pseudo-quasi overlap.

Proposição 4.2 *Sejam O_{PQ_1} e O_{PQ_2} pseudo-quasi overlaps e A uma função de média bivariada. Se $A(x, y) \in]0, 1[$ sempre que $x \in]0, 1[$ ou $y \in]0, 1[$, então $O_{PQ}(x, y) = A(O_{PQ_1}(x, y), O_{PQ_2}(x, y))$ é um pseudo-quasi overlap. Além disso,*

1. *Se $O_{PQ_2} = \overleftrightarrow{O}_{PQ_1}$ e A é simétrica, então O_{PQ} é um quasi-overlap;*
2. *Se A é um \wedge -morfismo e O_{PQ_1} e O_{PQ_2} são \wedge_l - (\wedge_r, \wedge) morfismo então O_{PQ} também é um \wedge_l - (\wedge_r, \wedge) morfismo;*
3. *Se A é um \vee -morfismo e O_{PQ_1} e O_{PQ_2} são \vee_L - (\vee_R, \vee) morfismo então O_{PQ} também é um \vee_L - (\vee_R, \vee) morfismo;*

4. Se O_{PQ_1} e O_{PQ_2} tem um elemento l -neutro (r -neutro, neutro) então O_{PQ} tem um elemento l -neutro (r -neutro, neutro);
5. Se O_{PQ_1} e O_{PQ_2} tem um elemento l -subneutro (r -subneutro, neutro) então O_{PQ} tem um elemento l -subneutro (r -subneutro, subneutro);
6. Se O_{PQ_1} e O_{PQ_2} são idempotentes então O_{PQ} também é idempotente;
7. Se O_{PQ_1} e O_{PQ_2} são sub-idempotentes então O_{PQ} também é sub-idempotente.

Prova: Sejam O_{PQ_1} e O_{PQ_2} *pseudo-quasi overlaps* e A uma função de média bivariada. Se $A(x, y) \in]0, 1[$ sempre que $x \in]0, 1[$ ou $y \in]0, 1[$, então O_{PQ} é claramente crescente. Além disso, já que A é uma função de média, por um lado, $O_{PQ}(x, y) = 0$ se, e somente se, $O_{PQ_1}(x, y) = 0$ e $O_{PQ_2}(x, y) = 0$ se, e somente se, $xy = 0$ e por outro lado, $O_{PQ}(x, y) = 1$ se, e somente se, $O_{PQ_1}(x, y) = 1$ e $O_{PQ_2}(x, y) = 1$ se, e somente se, $xy = 1$. Portanto, O_{PQ} é um *pseudo-quasi overlap*. Além disso,

1. Se $O_{PQ_2} = \overleftrightarrow{O}_{PQ_1}$ então para cada $x, y \in [0, 1]$, $O_{PQ}(x, y) = A(O_{PQ_1}(x, y), O_{PQ_1}(y, x)) = A(O_{PQ_1}(y, x), O_{PQ_1}(x, y)) = O_{PQ}(y, x)$. Portanto, O_{PQ} é simétrica e consequentemente é um *quasi-overlap*.
2. Se A é um \wedge -morfismo e O_{PQ_1} e O_{PQ_2} são \wedge_l -morfismos então $O_{PQ}(x \wedge y, z) = A(O_{PQ_1}(x \wedge y, z), O_{PQ_2}(x \wedge y, z)) = A(O_{PQ_1}(x, z) \wedge O_{PQ_1}(y, z), O_{PQ_2}(x, z) \wedge O_{PQ_2}(y, z)) = A(O_{PQ_1}(x, z), O_{PQ_2}(x, z)) \wedge A(O_{PQ_1}(y, z), O_{PQ_2}(y, z)) = O_{PQ}(x, z) \wedge O_{PQ}(y, z)$. O caso para \wedge_r -morfismos é análogo, e o caso para \wedge -morfismos é uma consequência de ambos.
3. Análogo ao item anterior.
4. Já que A é uma função de média então é idempotente. Então, se O_{PQ_1} e O_{PQ_2} tem 1 como um elemento l -neutro então $O_{PQ}(1, x) = A(O_{PQ_1}(1, x), O_{PQ_2}(1, x)) = A(x, x) = x$. O caso para elementos r -neutros é análogo e o caso de elemento neutro é uma consequência de ambos.
5. Se O_{PQ_1} e O_{PQ_2} tem 1 como elemento l -subneutro então $O_{PQ_1}(1, x) \leq x$ e $O_{PQ_2}(1, x) \leq x$. Então, já que A é uma função de média, $O_{PQ}(1, x) = A(O_{PQ_1}(1, x), O_{PQ_2}(1, x)) \leq x$. O caso para elementos r -subneutros é análogo e o caso para elemento subneutro é uma consequência de ambos.

6. Se O_{PQ_1} e O_{PQ_2} são idempotentes e já que A também é idempotente então $O_{PQ}(x, x) = A(O_{PQ_1}(x, x), O_{PQ_2}(x, x)) = A(x, x) = x$.
7. Se O_{PQ_1} e O_{PQ_2} são subidempotentes então $O_{PQ_1}(x, x) \leq x$ e $O_{PQ_2}(x, x) \leq x$. Então, já que A é idempotente e crescente, então $O_{PQ}(x, x) = A(O_{PQ_1}(x, x), O_{PQ_2}(x, x)) \leq A(x, x) = x$.

□

Corolário 4.1 *A soma convexa de dois pseudo-quasi overlaps O_{PQ_1} e O_{PQ_2} também é um pseudo-quasi overlap. Além disso, se O_{PQ_1} e O_{PQ_2} são \wedge_l -(\wedge_r , \wedge , \vee_L , \vee_R , \vee , l -neutro, r -neutro, neutro, l -subneutro, r -subneutro, subneutro, idempotente, subidempotente) morfismos, então O_{PQ} também é um \wedge_l -(\wedge_r , \wedge , \vee_L , \vee_R , \vee , l -neutro, r -neutro, neutro, l -subneutro, r -subneutro, subneutro, idempotente, subidempotente)*

Proposição 4.3 *Sejam O_{PQ_1} e O_{PQ_2} duas funções de pseudo-quasi overlap. Então $O_{PQ}(x, y) = O_{PQ_2}(O_{PQ_1}(x, y), O_{PQ_1}(y, x))$ é um quasi-overlap. Além disso, se O_{PQ_1} tem um elemento neutro e O_{PQ_2} é idempotente, então O_{PQ} tem um elemento neutro. Se O_{PQ_1} e O_{PQ_2} são idempotentes então O_{PQ} também é idempotente.*

Prova: Claramente, O_{PQ} é comutativa e as propriedades restantes listadas na Definição 2.12 seguem das mesmas propriedades que O_{PQ_1} e O_{PQ_2} . Por outro lado, se O_{PQ_1} tem elemento neutro e O_{PQ_2} é idempotente então

$$\begin{aligned} O_Q(x, 1) &= O_{PQ_2}(O_{PQ_1}(x, 1), O_{PQ_1}(1, x)) \\ &= O_{PQ_2}(x, x) \\ &= x \end{aligned}$$

e, se O_{PQ_1} e O_{PQ_2} são idempotentes então

$$\begin{aligned} O_Q(x, x) &= O_{PQ_2}(O_{PQ_1}(x, x), O_{PQ_1}(x, x)) \\ &= O_{PQ_2}(x, x) \\ &= x \end{aligned}$$

□

Proposição 4.4 *Sejam O_{PQ_1} e O_{PQ_2} duas funções de pseudo-quasi overlap. Então $O_Q(x, y) = O_{PQ_2}(O_{PQ_1}(x, 1), O_{PQ_1}(y, 1))$ também é um pseudo-quasi overlap. Além disso, se O_{PQ_1} e O_{PQ_2} tem um elemento r -neutro então O_Q tem elemento r -neutro. Ainda, se O_{PQ_2} for idempotente e O_{PQ_1} tiver elemento r -neutro, então O_Q é idempotente.*

Prova: Claramente, O_Q não é necessariamente comutativa e as propriedades restantes listadas na Definição 4.2 seguem das mesmas propriedades que O_{PQ_1} e O_{PQ_2} . Por outro lado, se O_{PQ_1} e O_{PQ_2} tem elemento r-neutro então

$$\begin{aligned} O_Q(x, 1) &= O_{PQ_2}(O_{PQ_1}(x, 1), O_{PQ_1}(1, 1)) \\ &= O_{PQ_2}(x, 1) \\ &= x \end{aligned}$$

e, se O_{PQ_2} é idempotente e O_{PQ_1} tem elemento r-neutro, então

$$\begin{aligned} O_Q(x, x) &= O_{PQ_2}(O_{PQ_1}(x, 1), O_{PQ_1}(x, 1)) \\ &= O_{PQ_2}(x, x) \\ &= x \end{aligned}$$

□

Proposição 4.5 *Se uma Cópula positiva é não simétrica, ela é um pseudo-quasi overlap, mas não é um quasi-overlap.*

Prova: Já que a cópula é não simétrica, ela não pode ser um *overlap* ou *quasi-overlap* já que eles são simétricos. Como toda Cópula tem 0 como elemento aniquilador, ela satisfaz automaticamente o lado esquerdo da propriedade 2 da Definição 2.11 e por ser positiva, satisfaz o lado direito da propriedade 2. Além disso, já que toda Cópula é crescente, satisfaz a propriedade 4. Finalmente, já que $C(x, 1) = x$ para todo $x \in [0, 1]$ então $C(1, 1) = 1$ e se $C(x, y) = 1$ então $C(x, 1) = 1$ e $C(y, 1) = 1$, e portanto $x = y = 1$. Portanto, satisfaz a propriedade 3. Desta forma, pela Definição 4.2, C é um *pseudo-quasi overlap*.

□

A Tabela 3 contém alguns exemplos de funções das classes previamente mencionadas, acompanhadas de uma nomenclatura, sua classificação e algumas propriedades da Definição 4.2. As tags MM, JM, N, rN, SN, I e SI nesta tabela correspondem as propriedades de \wedge -morfismo, \vee -morfismo, elemento neutro, elemento r-neutro, elemento subneutro, idempotente e subidempotente, respectivamente.

Teorema 4.3 *Seja $O_{PQ} : [0, 1]^2 \rightarrow [0, 1]$ uma função. O_{PQ} é uma pseudo-quasi overlap se, e somente se existem funções, $f, h : [0, 1]^2 \rightarrow [0, 1]$ tal que para cada $x, y \in [0, 1]$:*

1. f é crescente e h é decrescente;
2. $f(x, y) = 0$ se, e somente se $xy = 0$;
3. $h(x, y) = 0$ se, e somente se $xy = 1$;

Tabela 3: Definição de alguns *pseudo-quasi overlaps* e outras funções que serão utilizadas neste trabalho

Nomenclatura	Função	Classificação	Propriedades
O_{PO}^1	$O_{PO}^1(x, y) = \begin{cases} x^y & \text{se } y > 0 \\ 0 & \text{caso contrário} \end{cases}$	$\mathcal{PQO} - \mathcal{QO}$	rN, MM, JM
O_{PO}^2	$O_{PO}^2(x, y) = \begin{cases} \frac{2x^3y^2}{x^3+y^2} & \text{se } xy \neq 0 \\ 0 & \text{caso contrário} \end{cases}$	$\mathcal{PQO} - \mathcal{QO}$	SN, SI
O_{PO}^3	$O_{PO}^3(x, y) = \begin{cases} \frac{2x^e y^2}{x^e + y^2} & \text{se } xy \neq 0 \\ 0 & \text{caso contrário} \end{cases}$	$\mathcal{PQO} - \mathcal{QO}$	SN, SI
O_{PO}^4	$O_{PO}^4(x, y) = xy^2$	\mathcal{PO}	MM, JM, N, SI
O_Q^1	$O_Q^1(x, y) = \begin{cases} \frac{2xy}{x+y} & \text{se } x + y \neq 0 \\ 0 & \text{caso contrário} \end{cases}$	\mathcal{O}	I
O^1	x^2y^2	\mathcal{O}	MM, JM, SN, SI
C^1	$\frac{xy + \min(x, y)}{2}$	$\mathcal{O} \cap \mathcal{C}$	MM, N, SI
O^2	$\min(x\sqrt{y}, y\sqrt{x})$	$\mathcal{O} \cap \mathcal{C}$	MM, JM, N, SI
O_Q^2	$O_Q^2(x, y) = \begin{cases} 0.1 + 0.8x^2y^2 & \text{se } xy \notin \{0, 1\} \\ xy & \text{caso contrário} \end{cases}$	$\mathcal{QO} - (\mathcal{O} \cup \mathcal{C})$	MM, JM

$$4. O_{PQ}(x, y) = \frac{f(x, y)}{f(x, y) + h(x, y)}.$$

Prova: Análogo ao (BUSTINCE et al., 2010a, Teorema 5). □

4.3 Funções de (\cdot) -pré-agregação

Baseado na ideia de funções \vec{r} -crescentes, ordenadas direcionalmente (OD) \vec{r} -crescente, será introduzido um novo tipo de relaxamento da monotonicidade e com isso um relaxamento da noção de funções de agregação na mesma direção das funções de pré-agregação.

Definição 4.3 *Uma função de (\cdot) -pré-agregação é uma função $F : [0, 1]^n \rightarrow [0, 1]$ tal que existe um vetor positivo $\vec{r} = (r_1, \dots, r_n)$:*

1. F é (\vec{r}) -crescente, isto é, para cada $(x_1, \dots, x_n) \in [0, 1]^n$ e $c > 0$ tal que $(\vec{x}) + c\vec{r} = (x_{(1)} + r_1c, \dots, x_{(n)} + r_nc) \in L_n([0, 1])$, é verdade que $F((\vec{x})) \leq F((\vec{x}) + c\vec{r})$;
2. $F(0, \dots, 0) = 0$ e $F(1, \dots, 1) = 1$.

onde $L_n([0, 1]) = \{(a_1, \dots, a_n) \in [0, 1]^n : a_1 \leq \dots \leq a_n\}$.

Exemplo 4.4 *Seja F a função do Exemplo 2.1 e um vetor arbitrário $\vec{x} \in [0, 1]^n$, um vetor positivo real não nulo \vec{r} e $c > 0$. Se $(\vec{x}) + c\vec{r} = (x_{(1)} + r_1c, \dots, x_{(n)} + r_nc) \in L_n([0, 1])$*

então, já que, $k_1 = \lceil x_{(1)} + x_{(2)} + x_{(3)} \rceil \leq \lceil x_{(1)} + cr_1 + x_{(2)} + cr_2 + x_{(3)} + cr_3 \rceil = k_2$, temos que $x_{(k_1)} \leq x_{(k_2)}$ e portanto $x_{(k_1)} \leq x_{(k_1)} + cr_{k_1} \leq x_{(k_2)} + cr_{k_2}$. Então, $F(\vec{x}) \leq F(\vec{x} + \vec{c})$.

Desta forma, a partir dos exemplo anterior e do Exemplo 2.1, podemos concluir que ser SOD ou OD \vec{r} -crescente não é equivalente a ser (\vec{r}) -crescente.

4.4 Novas Generalizações da Integral de Choquet

Dada a importância das integrais de Choquet, diversas definições de generalização desta noção tem sido propostas (DIMURO et al., 2020a). Dentre estas variantes de integrais de Choquet, Lucca et al. (LUCCA et al., 2017) introduziu as Integrais de Choquet baseada em Copulas (CC-Integrals). Nesta tese, inicialmente este conceito será explorado do ponto de vista dos *quasi-overlaps*.

4.4.1 Integral de Choquet Utilizando *Quasi-Overlaps*

Esta nova Integral será chamada de Integral de Choquet Baseada em *Quasi-Overlaps* (CQO-Integral). A seguir, definiremos esta integral e exploraremos algumas de suas propriedades

Definição 4.4 *Seja $N = \{1, \dots, n\}$ um conjunto finito, $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy e $O : [0, 1]^2 \rightarrow [0, 1]$ um quasi-overlap. A CQO-Integral é definida como a função $\mathcal{C}_m^O : [0, 1]^n \rightarrow [0, n]$, para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$:*

$$\mathcal{C}_m^O(\vec{x}) = \sum_{i=1}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) \quad (4.1)$$

onde $(x_{(1)}, \dots, x_{(n)})$ é uma permutação crescente de \vec{x} . Por convenção, $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$.

Observação 4.3 \mathcal{C}_m^O é bem definida. De fato,

1. A escolha da permutação crescente, quando houver mais de uma, é irrelevante e ocorre só quando o vetor \vec{x} tiver elementos repetidos. Por simplicidade, supor que $(x_{(1)}, \dots, x_{(n)})$ e $(x_{[1]}, \dots, x_{[n]})$ são duas permutações crescentes de \vec{x} tais que $(k) = [k + 1]$ e $(k + 1) = [k]$ para algum $k \in \{1, \dots, n - 1\}$ e $(i) = [i]$ para todo $i \in$

$N - \{k, k + 1\}$. Portanto, $(*) x_{(k)} = x_{(k+1)}$, $(**) x_{(i)} = x_{[i]}$ para todo $i \in N$, e $(***) A_{(i)} \neq A_{[i]}$ se, e somente se $i = k + 1$. Seja $I = \{1, \dots, n\} - k + 1$. Logo, temos que

$$\begin{aligned}
\mathcal{C}_m^O(\vec{x}) &= \sum_{i=1}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) \\
&= \left(\sum_{i \in I} (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) \right) \\
&\quad + (O(x_{(k+1)}, \mathbf{m}(A_{(k+1)})) - O(x_{(k)}, \mathbf{m}(A_{(k+1)}))) \\
&= \sum_{i \in I} O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)})) \text{ por } (*) \\
&= \sum_{i \in I} O(x_{[i]}, \mathbf{m}(A_{[i]})) - O(x_{[i-1]}, \mathbf{m}(A_{[i]})) \text{ por } (**) \text{ e } (***) \\
&= \left(\sum_{i \in I} O(x_{[i]}, \mathbf{m}(A_{[i]})) - O(x_{[i-1]}, \mathbf{m}(A_{[i]})) \right) \\
&\quad + (O(x_{[k+1]}, \mathbf{m}(A_{[k+1]})) - O(x_{[k]}, \mathbf{m}(A_{[k+1]}))) \text{ por } (*) \text{ e } (**) \\
&= \sum_{i=1}^n (O(x_{[i]}, \mathbf{m}(A_{[i]})) - O(x_{[i-1]}, \mathbf{m}(A_{[i]})))
\end{aligned}$$

Por exemplo, se $\vec{x} = (x_1, \dots, x_4) = (0.5, 0.6, 0.5, 0.1)$ então (x_4, x_1, x_3, x_2) e (x_4, x_3, x_1, x_2) são ambas permutações crescentes de \vec{x} . Então, considerando a medida fuzzy relativa, por um lado temos que $\mathcal{C}_m^O(\vec{x}) = \sum_{i=1}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) = (O(0.1, 1) - 0) + (O(0.5, 0.6) - O(0.1, 0.6)) + (O(0.5, 0.5) - O(0.5, 0.5)) + (O(0.6, 0.2) - O(0.5, 0.2))$ e por outro lado, $\mathcal{C}_m^O(\vec{x}) = \sum_{i=1}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) = (O(0.1, 1) - 0) + (O(0.5, 0.6) - O(0.1, 0.6)) + (O(0.5, 0.3) - O(0.5, 0.3)) + (O(0.6, 0.2) - O(0.5, 0.2))$ e portanto, em ambos os casos, $\mathcal{C}_m^O(\vec{x}) = O(0.1, 1) + O(0.5, 0.6) - O(0.1, 0.6) + O(0.6, 0.2) - O(0.5, 0.2)$;

2. $\mathcal{C}_m^O(\vec{x}) \in [0, n]$ para todo $\vec{x} \in [0, 1]^n$. Já que O é crescente e $x_{(i-1)} \leq x_{(i)}$, então $O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)})) \in [0, 1]$ e portanto $\mathcal{C}_m^O(\vec{x}) \in [0, n]$.

Proposição 4.6 *Seja O um quasi-overlap e \mathbf{m} uma medida fuzzy. Então \mathcal{C}_m^O é idempotente se, e somente se O tem um elemento neutro.*

Prova: Seja $x \in [0, 1]$. Então $\mathcal{C}_m^O(x, \dots, x) = x$ se, e somente se $\sum_{i=1}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) = x$ se, e somente se $O(x_{(1)}, \mathbf{m}(A_{(1)})) - O(x_{(0)}, \mathbf{m}(A_{(1)})) + \sum_{i=2}^n (O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)}))) = x$ se, e somente se $O(x, 1) - O(0, 1) + 0 = x$ se, e somente se $O(x, 1) = x$. Logo, pela Observação 4.2, a proposição segue. \square

Proposição 4.7 *Seja O um quasi-overlap e \mathbf{m} uma medida fuzzy. Então $\mathcal{C}_m^O(0, \dots, 0) = 0$ e $\mathcal{C}_m^O(1, \dots, 1) = 1$.*

Prova: $\mathcal{C}_m^O(0, \dots, 0) = \sum_{i=1}^n (O(0, \mathbf{m}(A_{(i)})) - O(0, \mathbf{m}(A_{(i)}))) = 0$ e $\mathcal{C}_m^O(1, \dots, 1) = O(1, \mathbf{m}(A_{(1)})) - O(0, \mathbf{m}(A_{(1)})) + \sum_{i=2}^n (1, \mathbf{m}(A_{(i)})) - O(1, \mathbf{m}(A_{(i)})) = O(1, \mathbf{m}(A_{(1)})) = O(1, 1) = 1$. \square

Uma questão importante é estabelecer as condições suficientes e necessárias para $\text{Ran}(\mathcal{C}_m^O) \subseteq [0, 1]$. Em seguida, serão providas as condições suficientes

Proposição 4.8 *Seja O um quasi-overlap e \mathbf{m} uma medida fuzzy. Então $\text{Ran}(\mathcal{C}_m^O) \subseteq [0, 1]$ sempre que uma das seguintes condições é satisfeita*

1. \mathcal{C}_m^O é crescente;
2. O é 1-Lipschitz;
3. $\text{Ran}(\mathbf{m}) = \{0, 1\}$.

Prova: Seja $\vec{x} \in [0, 1]^n$. Já que $\mathcal{C}_m^O(\vec{x}) \in [0, n]$ então é suficiente provar que $\mathcal{C}_m^O(\vec{x}) \leq 1$.

1. Direto, já que $\mathcal{C}_m^O(\vec{x}) \leq \mathcal{C}_m^O(1, \dots, 1)$ e, pela Proposição 4.7, $\mathcal{C}_m^O(1, \dots, 1) = 1$;
2. Se O é 1-Lipschitz então $O(x_{(i)}, \mathbf{m}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}(A_{(i)})) \leq x_{(i)} - x_{(i-1)}$ e portanto, $\mathcal{C}_m^O(\vec{x}) \leq \sum_{i=1}^n x_{(i)} - x_{(i-1)} = x_{(n)} \leq 1$.
3. Se $\text{Ran}(\mathbf{m}) = \{0, 1\}$ então existe $k \in N$ tal que $\mathbf{m}(A_{(i)}) = 0$ para cada $(i) \in \{(k+1), \dots, (n)\}$ e $\mathbf{m}(A_{(i)}) = 1$ para todo $(i) \in \{(1), \dots, (k)\}$. Então, $\mathcal{C}_m^O(\vec{x}) = \sum_{i=1}^k (O(x_{(i)}, 1) - O(x_{(i-1)}, 1)) = O(x_{(k)}, 1) \leq 1$.

\square

Corolário 4.2 *Para qualquer Cópula C e medida fuzzy \mathbf{m} , $\text{Ran}(\mathcal{C}_m^C) \subseteq [0, 1]$.*

Teorema 4.4 *Seja O uma função de quasi-overlap e \mathbf{m} uma medida fuzzy. Então \mathcal{C}_m^O é uma função de agregação sempre que O é uma Cópula.*

Prova: Imediata a partir da Proposição 4.7 e (LUCCA et al., 2017, Prop. 5).

\square

Corolário 4.3 $\mathcal{C}_m^{O_B}$, $\mathcal{C}_m^{O_F}$ e $\mathcal{C}_m^{O_T}$ são funções de agregação.

Prova: Direto, já que, O_B e O_F são Cópulas.

\square

Proposição 4.9 *Seja O um quasi-overlap. Se O não é uma Cópula então existe uma medida fuzzy \mathbf{m} tal que \mathcal{C}_m^O não é uma função de agregação.*

Prova: Já que O não é uma Cópula, então existe $x, y, a, b \in [0, 1]$ tal que $x < y$, $a \leq b$ e $O(x, b) - O(x, a) > O(y, b) - O(y, a)$. Seja $\mathbf{m}_{a,b}$ a medida fuzzy

$$\mathbf{m}_{a,b}(X) = \begin{cases} 1 & \text{se } X = N \\ 0 & \text{se } X = \emptyset \\ b & \text{se } X = \{2, \dots, n\} \\ a & \text{caso contrário} \end{cases}$$

Sejam $\vec{x} = (0, x, x, 1, \dots, 1)$, $\vec{y} = (0, y, y, 1, \dots, 1)$ e $A_{(i)} = \{s \in N : x_s \geq x_{(i)}\}$ e $B_{(i)} = \{s \in N : y_s \geq y_{(i)}\}$ para todo $i \in N$. Claramente, temos que $\mathbf{m}_{a,b}(A_{(2)}) = \mathbf{m}_{a,b}(B_{(2)}) = b$ e $\mathbf{m}_{a,b}(A_{(i)}) = \mathbf{m}_{a,b}(B_{(i)}) = a$ para todo $i \in \{3, \dots, n\}$. Portanto,

$$\begin{aligned} \mathcal{C}_{\mathbf{m}_{a,b}}^O(\vec{x}) &= \sum_{i=1}^n (O(x_{(i)}, \mathbf{m}_{a,b}(A_{(i)})) - O(x_{(i-1)}, \mathbf{m}_{a,b}(A_{(i)}))) \\ &= O(0, 1) - O(0, 1) + O(x, \mathbf{m}(A_{(2)})) - O(0, \mathbf{m}(A_{(2)})) + O(x, \mathbf{m}(A_{(3)})) - O(x, \mathbf{m}(A_{(3)})) \\ &\quad + O(1, \mathbf{m}(A_{(4)})) - O(x, \mathbf{m}(A_{(4)})) + O(1, \mathbf{m}(A_{(5)})) - O(1, \mathbf{m}(A_{(5)})) + \dots + \\ &\quad O(1, \mathbf{m}(A_{(n)})) - O(1, \mathbf{m}(A_{(n)})) \\ &= O(x, \mathbf{m}(A_{(2)})) + O(1, \mathbf{m}(A_{(4)})) - O(x, \mathbf{m}(A_{(4)})) \\ &= O(x, b) + O(1, a) - O(x, a) \\ &> O(y, b) + O(1, a) - O(y, a) \\ &= O(0, 1) - O(0, 1) + O(y, \mathbf{m}(B_{(2)})) - O(0, \mathbf{m}(B_{(2)})) + O(y, \mathbf{m}(B_{(3)})) - O(y, \mathbf{m}(B_{(3)})) \\ &\quad + O(1, \mathbf{m}(B_{(4)})) - O(y, \mathbf{m}(B_{(4)})) + O(1, \mathbf{m}(B_{(5)})) - O(1, \mathbf{m}(B_{(5)})) + \dots + \\ &\quad O(1, \mathbf{m}(B_{(n)})) - O(1, \mathbf{m}(B_{(n)})) \\ &= \sum_{i=1}^n (O(y_{(i)}, \mathbf{m}(B_{(i)})) - O(y_{(i-1)}, \mathbf{m}(B_{(i)}))) \\ &= \mathcal{C}_m^O(\vec{y}) \end{aligned}$$

□

A Proposição 4.9 não exclui a possibilidade de CQO-integrais serem funções de agregação quando forem geradas a partir de pseudo-quase overlaps que não são cópulas. De fato, para

$$\mathbf{m}_\top(X) = \begin{cases} 0 & \text{se } X = \emptyset \\ 1 & \text{caso contrário} \end{cases}$$

temos que $\mathcal{C}_{\mathbf{m}_\top}^O = \max$ sempre que O é um quasi-overlap com 1 como elemento neutro,

como por exemplo, O_σ e O_C .

Proposição 4.10 *Seja O um quasi-overlap e \mathbf{m} uma medida fuzzy. Então \mathcal{C}_m^O é uma função de pré-agregação se para cada $x, y, z \in [0, 1]$ e $c > 0$, $O(y, z) - O(x, z) \leq O(y + c, z) - O(x + c, z)$ sempre que $x \leq y$ e $x + c, y + c \in [0, 1]$.*

Prova: Pode-se provar que \mathcal{C}_m^O é $\vec{1} = (1, \dots, 1)$ -crescente.

Seja $\vec{x} \in [0, 1]^n$ e $c \in \mathbb{R}^+$ tal que $\vec{y} = (x_1 + c, \dots, x_n + c) \in [0, 1]^n$. Então é imediato que $A_{(i)} = \{s : x_s \geq x_{(i)}\} = \{s : y_s \geq y_{(i)}\} = B_{(i)}$. Portanto,

$$\begin{aligned} \mathcal{C}_m^O(\vec{x}) &= \sum_{i=1}^n (O(x_{(i)}, m(A_{(i)})) - O(x_{(i-1)}, m(A_{(i)}))) \\ &\leq \sum_{i=1}^n (O(x_{(i)} + c, m(A_{(i)})) - O(x_{(i-1)} + c, m(A_{(i)}))) \\ &= \mathcal{C}_m^O(\vec{y}) \end{aligned}$$

□

Corolário 4.4 $\mathcal{C}_m^{O_\sigma}$, $\mathcal{C}_m^{O_C}$, $\mathcal{C}_m^{O_T}$ e $\mathcal{C}_m^{O_2}$ são funções de pré-agregação.

Prova: Direto, a partir da proposição 4.10. □

4.4.2 Integral de Choquet Utilizando *Pseudo Overlaps*

Agora definiremos uma extensão natural da definição anterior da CQO-Integral, substituindo o *quasi-overlap* por um *pseudo overlap*.

Definição 4.5 *Seja $PO : [0, 1]^2 \rightarrow [0, 1]$ um pseudo overlap e $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy. A pseudo integral de Choquet é definida como a função $\mathcal{PC}_m^{PO} : [0, 1]^n \rightarrow [0, 1]$ dada para cada $\vec{x} = (x_1, \dots, x_n)$ por*

$$\mathcal{PC}_m^{PO}(\vec{x}) = \sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)}))$$

onde $(x_{(1)}, \dots, x_{(n)})$ é uma permutação crescente de \vec{x} , $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$.

O Lema 4.2 mostra que se há duas ou mais permutações crescentes para uma entrada \vec{x} , a escolha de qual delas usar na computação de $\mathcal{PC}_m^{PO}(\vec{x})$ é irrelevante. Note que há mais de uma permutação crescente para um vetor \vec{x} se, e somente se o vetor tem elementos repetidos.

Lema 4.2 *Seja PO uma função pseudo overlap e \mathbf{m} uma medida fuzzy. Se $(x_{(1)}, \dots, x_{(n)})$ e $(x_{[1]}, \dots, x_{[n]})$ são duas permutações diferentes e crescentes de um vetor $\vec{x} \in [0, 1]^n$ então*

$$\sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) = \sum_{i=1}^n PO(x_{[i]} - x_{[i-1]}, \mathbf{m}(A_{[i]}))$$

Prova: Sem perda de generalidade assumamos que as duas permutações crescentes de \vec{x} , $(x_{(1)}, \dots, x_{(n)})$ e $(x_{[1]}, \dots, x_{[n]})$, são tais que $(k) = [k + 1]$ e $(k + 1) = [k]$ para um $k \in \{1, \dots, n - 1\}$, $(i) = [i]$ para todo $i \notin \{k, k + 1\}$ e $(*)$ $x_{(k)} = x_{(k+1)}$. Portanto, $(**)$ $x_{(i)} = x_{[i]}$ para todo $i \in N$ e $(***)$ $A_{(i)} \neq A_{[i]}$ se, e somente se $i = k + 1$. Seja $I = \{1, \dots, n\} - \{k + 1\}$. Então,

$$\begin{aligned} \sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) &= \left(\sum_{i \in I} PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \right) + PO(x_{(k+1)} - x_{(k)}, \mathbf{m}(A_{(k+1)})) \\ &= \sum_{i \in I} PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \text{ por } (*) \text{ e } \mathbf{(PO1)} \\ &= \sum_{i \in I} PO(x_{[i]} - x_{[i-1]}, \mathbf{m}(A_{[i]})) \text{ por } (**) \text{ e } (***) \\ &= \left(\sum_{i \in I} PO(x_{[i]} - x_{[i-1]}, \mathbf{m}(A_{[i]})) \right) + PO(x_{[k+1]} - x_{[k]}, \mathbf{m}(A_{[k+1]})) \\ &\quad \text{por } (*), (**) \text{ e } \mathbf{(PO1)} \\ &= \sum_{i=1}^n PO(x_{[i]} - x_{[i-1]}, \mathbf{m}(A_{[i]})) \end{aligned}$$

□

Proposição 4.11 *Seja PO uma função pseudo overlap e \mathbf{m} uma medida fuzzy. Se PO tem um elemento r -subneutro então $\mathcal{PC}_{\mathbf{m}}^{PO}$ está bem definida.*

Prova: Pelo Lema 4.2 a computação de $\mathcal{PC}_{\mathbf{m}}^{PO}$ para um $\vec{x} \in [0, 1]^n$ não depende da permutação crescente escolhida. Assim, so resta provar que $\mathcal{PC}_{\mathbf{m}}^{PO}(\vec{x}) \in [0, 1]$ para todo $\vec{x} \in [0, 1]^n$. De fato, como PO tem um elemento r -subneutro e ser crescente, segue que $\sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \leq \sum_{i=1}^n x_{(i)} - x_{(i-1)} = x_{(n)} \leq 1$. Além disso, como $PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \geq 0$ para todo $i \in N$ e $\vec{x} \in [0, 1]^n$, então $\mathcal{PC}_{\mathbf{m}}^{PO}(\vec{x}) \in [0, 1]$. □

Proposição 4.12 *Seja PO uma função pseudo overlap e \mathbf{m} uma medida fuzzy. PO é r -subneutro se, e somente se $\mathcal{PC}_{\mathbf{m}}^{PO}$ é sub-idempotente. Além disso, 1 é um elemento r -neutro de PO se, e somente se $\mathcal{PC}_{\mathbf{m}}^{PO}$ é idempotente.*

Prova: Trivialmente,

$$\mathcal{PC}_{\mathbf{m}}^{PO}(x, \dots, x) = PO(x, 1) \tag{4.2}$$

De fato, $\mathcal{PC}_{\mathbf{m}}^{PO}(x, \dots, x) = \sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) = PO(x_{(1)}, \mathbf{m}(N)) + \sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) = PO(x, 1) + \sum_{i=1}^n PO(x - x, \mathbf{m}(A_{(i)})) = PO(x, 1)$ para todo $x \in [0, 1]$.

Portanto, pela Equação (4.2), PO é r -subneutro se, e somente se \mathcal{PC}_m^{PO} é sub-idempotente. Além disso, novamente pela Equação (4.2), \mathcal{PC}_m^{PO} é idempotente se, e somente se 1 é um elemento r -neutro de PO . \square

Teorema 4.5 *Seja $PO : [0, 1]^2 \rightarrow [0, 1]$ um pseudo overlap r -subneutro e $\mathbf{m} : \mathcal{P}(N) \rightarrow [0, 1]$ uma medida fuzzy. Então \mathcal{PC}_m^{PO} é uma (\cdot) -função de pré-agregação sub-idempotente.*

Prova: Pela Proposição 4.11 \mathcal{PC}_m^{PO} está bem definido e é sub-idempotente pela Proposição 4.12. Por outro lado, trivialmente, $\mathcal{PC}_m^{PO}(\vec{0}) = \sum_{i=1}^n PO(0, \mathbf{m}(A_{(i)})) = 0$ e $\mathcal{PC}_m^{PO}(\vec{1}) = PO(1-0, \mathbf{m}(A_{(1)})) + \sum_{i=2}^n PO(0, A_{(i)}) = PO(1, 1) = 1$. Finalmente, provaremos que \mathcal{PC}_m^{PO} é (\vec{r}) -crescente para o vetor positivo $\vec{r} = (0, \dots, 0, 1)$. Então,

$$\begin{aligned} \mathcal{PC}_m^{PO}((\vec{x})) &= \sum_{i=1}^n PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \\ &= PO(x_{(n)} - x_{(n-1)}, \mathbf{m}(A_{(n)})) + \sum_{i=1}^{n-1} PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \\ &\leq PO((x_{(n)} + c) - x_{(n-1)}, \mathbf{m}(A_{(n)})) + \sum_{i=1}^{n-1} PO(x_{(i)} - x_{(i-1)}, \mathbf{m}(A_{(i)})) \\ &= \mathcal{PC}_m^{PO}((\vec{x}) + c\vec{r}). \end{aligned}$$

\square

4.4.3 Integral de Choquet Absoluta

Lucca et al. (2019) introduziu uma nova generalização das integrais de Choquet, denominada de $C_{F_1 F_2}$, que são uma família de funções de agregação que utilizam duas funções de fusão F_1 e F_2 , mostradas na Definição 2.18. No entanto, essa variante tem duas principais limitações que foram abordadas nessa tese:

- A condição de dominação de F_1 sobre F_2 é um limitante importante, pois nem sempre é fácil provar a dominância de uma função sobre outra, e, além disso, existem muitos pares de funções F_1 e F_2 em que nem F_1 domina F_2 , nem o contrário. O que limita as opções de escolha para utilização na integral $C_{F_1 F_2}$.
- A mesma medida fuzzy \mathbf{m} deve ser usada na integral $C_{F_1 F_2}$, o que também não permite o uso de medidas fuzzy distintas para tentar melhorar a acurácia da classificação.

Neste sentido, nesta tese também serão introduzidas as Integrais de Choquet Absolutas (AC-Integral, do inglês *Absolute Choquet*), com o objetivo de tratar os pontos citados anteriormente. Aqui, serão utilizadas duas funções de agregação F_1 e F_2 , e serão verificadas as condições necessárias para, que a AC-Integral seja uma função de agregação, ou,

pelo menos, algum tipo de pré-agregação. Além disso, o módulo será incluído na integral para evitar valores negativos.

Definição 4.6 *Seja $N = \{1, \dots, n\}$ um conjunto finito, $\mathbf{m}_1, \mathbf{m}_2 : \mathcal{P}(N) \rightarrow [0, 1]$ duas medidas fuzzy e $F_1, F_2 : [0, 1]^2 \rightarrow [0, 1]$ duas funções de agregação. Uma $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ -Integral é definida como a função $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2} : [0, 1]^n \rightarrow [0, n]$, dada para todo $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$ por:*

$$AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(\vec{x}) = \sum_{i=1}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \quad (4.3)$$

onde $(x_{(1)}, \dots, x_{(n)})$ é a permutação crescente de $\vec{x} = (x_1, \dots, x_n)$ tal que se $x_{(i)} = x_{(i+1)}$, para algum $i \in N$, então $(i) < (i+1)$. Por convenção $x_{(0)} = 0$ e $A_{(i)} = \{(i), \dots, (n)\}$.

Observação 4.4 *Toda CQO-Integral é uma AC-Integral*

Observação 4.5 *As AC-Integrais são bem definidas. De fato, já que $0 \leq |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \leq 1$ então $0 \leq \sum_{i=1}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \leq n$. Portanto, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(\vec{x}) \in [0, n]$. Além disso, observe que para qualquer $\vec{x} \in [0, 1]^n$ só existe uma única permutação crescente $(x_{(1)}, \dots, x_{(n)})$ de \vec{x} tal que para todo $i \in N$, $x_{(i)} = x_{(i+1)} \Rightarrow (i) < (i+1)$ a qual denominaremos de permutação crescente absoluta de \vec{x} .*

Proposição 4.13 *Sejam F_1 e F_2 funções de agregação e \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy. Então,*

1. *Se 0 é um aniquilador à esquerda de F_1 e F_2 então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(0, \dots, 0) = 0$;*
2. *Se $F_1(1, y) = F_2(1, z)$ para todo $y, z \in [0, 1]$ e \mathbf{m}_1 e \mathbf{m}_2 são estritamente limitadas então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = 1$;*
3. *Se 1 é um aniquilador à esquerda de F_1 e F_2 então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = 1$;*
4. *Se 1 é um elemento l-neutro de F_1 e F_2 e $\mathbf{m}_1 = \mathbf{m}_2$ então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = 1$;*
5. *Se $F_1(x, 1) \geq x$ para todo $x \in [0, 1]$ então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2} \geq \min$.*

Prova:

1. Se 0 é um aniquilador à esquerda de F_1 e F_2 então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(0, \dots, 0) = \sum_{i=1}^n |F_1(0, \mathbf{m}_1(A_{(i)})) - F_2(0, \mathbf{m}_2(A_{(i)}))| = 0$.
2. Se $F_1(1, y) = F_2(1, z)$ para cada $y, z \in [0, 1]$ e \mathbf{m}_1 e \mathbf{m}_2 são estritamente limitados então para todo $i = 2, \dots, n$, $\mathbf{m}_1(A_{(i)}) \in]0, 1[$ e portanto $F_1(1, \mathbf{m}_1(A_{(i)})) = F_2(1, \mathbf{m}_2(A_{(i)}))$. Então, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = |F_1(1, 1) - F_2(0, 1)| + \sum_{i=2}^n |F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)}))| = |F_1(1, 1) - F_2(0, 1)| = 1$.
3. Se 1 é um aniquilador à esquerda de F_1 e F_2 então $F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)})) = 0$ para todo $i = 2, \dots, n$ e portanto $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = |F_1(1, 1) - F_2(0, 1)| + \sum_{i=2}^n |F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)}))| = |F_1(1, 1) - F_2(0, 1)| = 1$.
4. Se 1 é um elemento l-neutro de F_1 e F_2 e $\mathbf{m}_1 = \mathbf{m}_2$ então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = |F_1(1, 1) - F_2(0, 1)| + \sum_{i=2}^n |F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)}))| = |F_1(1, 1) - F_2(0, 1)| + \sum_{i=2}^n |\mathbf{m}_1(A_{(i)}) - \mathbf{m}_2(A_{(i)})| = 1$;
5. Se $F_1(x, 1) \geq x$ para todo $x \in [0, 1]$ então

$$\begin{aligned}
AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(x_1, \dots, x_n) &= |F_1(x_{(1)}, \mathbf{m}_1(A_{(1)})) - F_2(0, \mathbf{m}_2(A_{(1)}))| + \\
&\quad \sum_{i=2}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\
&= |F_1(x_{(1)}, 1) - 0| + \sum_{i=2}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\
&\geq x_{(1)} + \sum_{i=2}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\
&\geq x_{(1)} = \min(x_1, \dots, x_n)
\end{aligned}$$

□

Corolário 4.5 *Sejam F_1 e F_2 pseudo-quasi overlaps com 1 como elemento neutro e \mathbf{m} uma medida fuzzy. Então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(0, \dots, 0) = 0$ e $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = 1$. Além disso, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2} \geq \min$.*

Prova: Note que se F_1 e F_2 tem 1 como elemento neutro, o 0 é um elemento aniquilador a esquerda. Portanto, é imediato a partir da Proposição 4.13. □

Proposição 4.14 *Sejam F_1 e F_2 funções de agregação e \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy. Então $\text{Ran}(AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}) \subseteq [0, 1]$ se uma das seguintes condições é verdadeira:*

1. $F_1 = F_2$, $\mathbf{m}_1 = \mathbf{m}_2$ e a primeira seção de F_1 é 1-Lipstchitz;

2. $|F_1(y, \mathbf{m}_1(X)) - F_2(x, \mathbf{m}_2(X))| \leq y - x$ sempre que $x \leq y$ e $X \subseteq N$;
3. $\mathbf{m}_1 = \mathbf{m}_2$, $Ran(\mathbf{m}_1) = \{0, 1\}$, 0 é um aniquilador à direita e 1 é um elemento r -neutro de F_1 e F_2 .

Prova:

1. Já que a primeira seção de F_1 é 1-Lipstchitz e $x_{(i-1)} \leq x_{(i)}$ então $F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_1(x_{(i-1)}, \mathbf{m}_1(A_{(i)})) \leq x_{(i)} - x_{(i-1)}$ e portanto,

$$\begin{aligned} AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(x_1, \dots, x_n) &= \sum_{i=1}^n (F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_1(x_{(i-1)}, \mathbf{m}_1(A_{(i)}))) \\ &\leq \sum_{i=1}^n x_{(i)} - x_{(i-1)} \\ &= x_{(n)} \\ &\leq 1. \end{aligned}$$

2. Se $|F_1(y, \mathbf{m}_1(X)) - F_2(x, \mathbf{m}_2(X))| \leq y - x$ sempre que $x \leq y$ e $X \subseteq N$, então $|F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \leq x_{(i)} - x_{(i-1)}$ para todo $\vec{x} \in [0, 1]^n$. Então,

$$\begin{aligned} AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(x_1, \dots, x_n) &= \sum_{i=1}^n (F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))) \\ &\leq \sum_{i=1}^n x_{(i)} - x_{(i-1)} \\ &= x_{(n)} \\ &\leq 1. \end{aligned}$$

3. Se $\mathbf{m}_1 = \mathbf{m}_2$, $Ran(\mathbf{m}_1) = \{0, 1\}$, 0 é um elemento aniquilador à direita e 1 é um elemento r -neutro de F_1 e F_2 então para todo $\vec{x} \in [0, 1]^n$ existe $k \in N$ tal que $\mathbf{m}_1(A_{(i)}) = 0$ para todo $(i) \in \{(k_1), \dots, (n)\}$. Então,

$$\begin{aligned} AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(x_1, \dots, x_n) &= \sum_{i=1}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_1(A_{(i)}))| \\ &= \sum_{i=1}^k |F_1(x_{(i)}, 1) - F_2(x_{(i-1)}, 1)| \\ &= \sum_{i=1}^k (x_{(i)} - x_{(i-1)}) \\ &= x_{(k)} \\ &\leq 1. \end{aligned}$$

□

Proposição 4.15 *Sejam F_1 e F_2 pseudo-quasi overlaps, e \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy. Se 1 é um elemento r -neutro de F_1 , 0 é um aniquilador à esquerda de F_1 e F_2 , e $F_1(x, y) = F_2(x, z)$ para todo $x, y, z \in (0, 1)$, então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}$ é idempotente.*

Prova: Já que $F_1(x, \mathbf{m}_1(A_{(i)})) = F_2(x, \mathbf{m}_2(A_{(i)}))$ para todo $i = 2, \dots, n$, então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(x, \dots, x) = |F_1(x, 1) - F_2(0, 1)| + \sum_{i=2}^n |F_1(x, \mathbf{m}_1(A_{(i)})) - F_2(x, \mathbf{m}_2(A_{(i)}))| = |F_1(x, 1) - F_2(0, 1)| = x$. \square

Proposição 4.16 *Sejam \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy F_1 e F_2 funções de agregação tal que, para cada $x \leq y \in [0, 1]$, $X \subseteq N$ e $c \leq \min(1 - x, 1 - y)$,*

$$|F_1(x, \mathbf{m}_1(X)) - F_2(y, \mathbf{m}_2(X))| \leq |F_1(x + c, \mathbf{m}_1(X)) - F_2(y + c, \mathbf{m}_2(X))|. \quad (4.4)$$

$AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é (\vec{r}) -crescente para $\vec{r} = (1, \dots, 1)$.

Prova: Seja $\vec{r} = (1, \dots, 1)$, $c > 0$ e $\vec{x} = (x_1, \dots, x_n) \in [0, 1]^n$ tal que $x_{(n)} + c \leq 1$. Considere $\vec{y} = (\vec{x}) + c\vec{r} = (x_{(1)} + c, \dots, x_{(n)} + c)$. Note que $y_{(i)} = x_{(i)} + c = y_i$ e portanto $A_{(i)}$, com respeito a (\vec{x}) e \vec{y} são os mesmos, ou seja, $A_{(i)} = \{i, i + 1, \dots, n\}$. Então, pela Equação (4.4),

$$\begin{aligned} AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(\vec{y}) &= \sum_{i=1}^n |F_1(y_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(y_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\ &= F_1(x_{(1)} + c, 1) + \sum_{i=2}^n |F_1(x_{(i)} + c, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)} + c, \mathbf{m}_2(A_{(i)}))| \\ &\geq F_1(x_{(1)}, 1) + \sum_{i=2}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\ &= AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}((\vec{x})) \end{aligned}$$

Portanto, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é (\vec{r}) -crescente. \square

Proposição 4.17 *Sejam \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy tal que $\mathbf{m}_1(\{n\}) = \mathbf{m}_2(\{n\}) = a$ e F_1 e F_2 sejam funções de agregação tal que, para cada $x \leq y \in [0, 1]$, $X \subseteq N$ e $c \leq \min(1 - x, 1 - y)$,*

$$|F_1(x, a) - F_2(y, a)| \leq |F_1(x + c, a) - F_2(y, a)|. \quad (4.5)$$

$AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é (\vec{r}) -crescente para $\vec{r} = (0, \dots, 0, 1)$.

Prova: Seja $\vec{r} = (0, \dots, 0, 1)$, $\vec{x} = (x_1, \dots, x_n)$, $c \leq 1 - x_{(n)}$ e $\vec{y} = (\vec{x}) + c\vec{r} = (x_{(1)}, \dots, x_{(n-1)}, x_{(n)} + c)$. Então $y_{(i)} = y_i$ e portanto o $A_{(i)}$ associado a \vec{y} é igual ao $A_{(i)}$ associado a (\vec{x}) , ou seja $A_{(i)} = \{i, \dots, n\}$.

$$\begin{aligned}
AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(\vec{y}) &= \sum_{i=1}^n |F_1(y_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(y_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\
&= \left(\sum_{i=1}^{n-1} |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \right) + |F_1(x_{(n)} + c, a) - F_2(x_{(n-1)}, a)| \\
&\geq \left(\sum_{i=1}^{n-1} |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \right) + \\
&\quad |F_1(x_{(n)}, a) - F_2(x_{(n-1)}, a)| \text{ by Eq. (4.5)} \\
&= \sum_{i=1}^n |F_1(x_{(i)}, \mathbf{m}_1(A_{(i)})) - F_2(x_{(i-1)}, \mathbf{m}_2(A_{(i)}))| \\
&= AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(\vec{x}).
\end{aligned}$$

Então, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é (\vec{r}) -crescente. \square

Teorema 4.6 *Sejam \mathbf{m}_1 e \mathbf{m}_2 medidas fuzzy e F_1 e F_2 sejam funções de agregação tal que, para cada $x \leq y \in [0, 1]$, $X \subseteq N$ e $c \leq \min(1 - x, 1 - y)$ temos que*

$$|F_1(y, \mathbf{m}_1(X)) - F_2(x, \mathbf{m}_2(X))| \leq y - x \quad (4.6)$$

Se $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}$ é (\cdot) -crescente e 0 é um elemento aniquilador a esquerda de F_1 e F_2 então $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}$ é uma (\cdot) -pre-agregação.

Prova: Já que, pela Proposição 4.13, temos que $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}(0, \dots, 0) = 0$, pela hipótese temos que $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}$ é (\cdot) -crescente e pela Proposição 4.14 que $Ran(AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}) \subseteq [0, 1]$ então só resta provar que $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}(1, \dots, 1) = 1$. De fato, pela Equação (4.6),

$$\begin{aligned}
AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) &= F_1(1, 1) - F_2(0, 1) + \sum_{i=2}^n |F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)}))| \\
&= 1 + \sum_{i=2}^n |F_1(1, \mathbf{m}_1(A_{(i)})) - F_2(1, \mathbf{m}_2(A_{(i)}))| \\
&\leq 1 + \sum_{i=2}^n (x_{(i)} - x_{(i-1)}) \\
&= 1.
\end{aligned}$$

Então, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}(1, \dots, 1) = 1$. Desta formar, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_2 F_1}$ é uma (\cdot) -pre-agregação. \square

Exemplo 4.5 *Seja \mathbf{m}_1 uma medida fuzzy e $\mathbf{m}_2(X) = \sqrt{\mathbf{m}_1(X)}$ para cada $X \subseteq P(\mathbb{N})$. Então \mathbf{m}_2 também é uma medida fuzzy. Seja $F_1(x, y) = xy$ e $F_2(x, y) = xy^2$. Então F_1 e F_2 são pseudo-quasi overlaps tal que $F_1 \geq F_2$. Se $x \geq y$, $c \leq 1 - y$ e $a = \mathbf{m}_1(X)$ então $F_2(y, \sqrt{a}) = ya \leq xa = F_1(x, a)$ e $F_2(y + c, \sqrt{a}) = (y + c)a \leq (x + c)a = F_1(x + c, a)$. Portanto, $|F_1(x, \mathbf{m}_1(X)) - F_2(y, \mathbf{m}_2(X))| = F_1(x, a) - F_2(y, a^2) = xa - ya = (x + c)a - (y + c)a = |F_1(x + c, \mathbf{m}_1(X)) - F_2(y + c, \mathbf{m}_2(X))|$. Então, $(F_1, F_2, \mathbf{m}_1, \mathbf{m}_2)$ satisfazem a*

Equação (4.4). Além disso, já que $|F_1(y, \mathbf{m}_1(X)) - F_2(x, \mathbf{m}_2(X))| = F_1(x, a) - F_2(y, a^2) = xa - ya \leq y - x$ e portanto, $(F_1, F_2, \mathbf{m}_1, \mathbf{m}_2)$ satisfazem a Equação (4.6). Portanto, já que 0 é um elemento aniquilador a esquerda de F_1 e F_2 então, pelo Teorema 4.6, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é uma (\cdot) -pre-agregação.

Teorema 4.7 *Sejam \mathbf{m}'_1 and \mathbf{m}'_2 duas medidas fuzzy arbitrárias e F_1 e F_2 dois pseudo-quasi-overlaps. Então, existem medidas fuzzy \mathbf{m}_1 e \mathbf{m}_2 tal que $\mathbf{m}_1(\{n\}) = \mathbf{m}_2(\{n\}) = 0$, $\mathbf{m}_1(X) = \mathbf{m}'_1(X)$ e $\mathbf{m}_2(X) = \mathbf{m}'_2(X)$ para cada $X \neq \{n\}$ e $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é uma (\cdot) -pre-agregação.*

Prova: Sejam \mathbf{m}'_1 e \mathbf{m}'_2 duas medidas fuzzy arbitrárias. Então, as funções

$$\mathbf{m}_i(X) = \begin{cases} \mathbf{m}'_i(X) & , \text{ se } X \neq \{n\} \\ 0 & , \text{ caso contrário} \end{cases}$$

para $i = 1$ e $i = 2$, claramente são medidas fuzzy tal que $\mathbf{m}_1(\{n\}) = 0 = \mathbf{m}_2(\{n\})$. Então para qualquer pseudo-quasi-overlaps F_1 e F_2 , e $x \leq y$, temos que $|F_1(y, \mathbf{m}_1(\{n\})) - F_2(x, \mathbf{m}_2(\{n\}))| = |F_1(x, 0) - F_2(y, 0)| = 0$. Portanto, $(F_1, F_2, \mathbf{m}_1, \mathbf{m}_2)$ satisfazem as Equações (4.5) e (4.6). Então, já que 0 é um elemento aniquilador de F_1 e F_2 e pelo Teorema 4.6, $AC_{\mathbf{m}_1, \mathbf{m}_2}^{F_1 F_2}$ é uma (\cdot) -pre-agregação. \square

4.5 Considerações Finais

Neste capítulo, foram apresentadas as principais contribuições teóricas deste trabalho. Foi iniciada a discussão apresentando duas generalização das Integrais de Choquet, utilizando *quasi-overlaps* e *pseudo overlaps*. Em seguida, foi apresentada a principal contribuição deste trabalho: as Integrais de Choquet absolutas. Também foram definida duas novas classes de funções de agregação denominadas *pseudo overlaps* e *pseudo-quasi overlaps*.

5 Metodologia Proposta e Configuração Experimental

Neste capítulo, serão mostradas as contribuições metodológicas propostas nesta tese. Além disso, serão estabelecidos os parâmetros e configurações utilizadas para os experimentos realizados no capítulo seguinte. Também nesse capítulo será estabelecido qual o papel que as generalizações da integral de Choquet irão desempenhar no comitê de classificadores.

5.1 Arquitetura do Comitê de Classificadores Proposta

Nesta seção, iremos descrever a arquitetura que será proposta com relação ao Comitê de Classificadores que será utilizado, bem como definiremos como será o funcionamento do sistema, através de algoritmos e esquematizações.

Neste trabalho, não iremos propor mudanças na etapa de treinamento dos classificadores. Portanto, para as definições seguintes, considere que todos os classificadores foram previamente treinados e já estão prontos para receber elementos para serem classificados. O algoritmo inicia recebendo um conjunto de dados e um conjunto de classificadores, representado pelo Algoritmo 1

Algoritmo 1: Pseudo-código para classificação de elementos ω

```

1 Entrada:
2  $\Omega \leftarrow X_1, \dots, X_q$  - Um conjunto de dados com  $q$  elementos
3  $\mathcal{C} \leftarrow C_1, \dots, C_n$  - Um conjunto com  $n$  classificadores
4 Saída:
5  $F$  - um vetor com as classificações finais dos elementos
6 classify-dataset( $\Omega, C$ ):
7 for  $j = 1$  to  $q$  do
8   |  $F_j \leftarrow$  classify-instance-ac-integral( $\Omega_j, C$ )
9 end
10 return  $F$ 

```

Inicialmente, definimos um conjunto de classificadores $\mathfrak{C} = \{C_1, \dots, C_n\}$ que irão compor o comitê definido. Não haverá restrição sobre quais classificadores pertencem ao comitê, portanto pode-se considerar este como um comitê heterogêneo. Agora, suponha um elemento $X = (x_1, \dots, x_k)$ de k atributos, que irá ser classificado, em um problema onde há p classes possíveis. As etapas seguintes estão ilustradas no Algoritmo 2.

Algoritmo 2: Pseudo-código para o modelo proposto

```

1 Entrada:
2  $\mathfrak{C} \leftarrow C_1, \dots, C_n$  - Um conjunto com  $n$  classificadores
3  $X \leftarrow x_1, \dots, x_k$  - Um elemento com  $k$  atributos
4 Saída:
5  $class(x)$  - a classificação final do elemento
6 classify-instance-ac-integral( $\mathfrak{C}, \mathbf{x}$ ):
7  $V \leftarrow$  é uma matriz de dimensões  $n \times p$  inicialmente vazia
8  $Y \leftarrow$  é um vetor de  $p$  posições inicialmente vazio
9 for  $j = 1$  to  $n$  do
10 |  $V[j] \leftarrow \mathbf{classify}_{C_j}(\mathbf{x})$ 
11 end
12 for  $w = 1$  to  $p$  do
13 |  $Y[w] \leftarrow \mathbf{aggregate-ac-integral}(V[1][w], \dots, V[n][w])$ 
14 end
15 return  $\mathbf{argmax}(Y)$ 

```

Como foi anteriormente mencionado, comitês de classificadores possuem duas camadas: a camada de classificação e a camada de combinação. Na etapa de classificação, inicialmente cada classificador irá processar o elemento a ser classificado (representado na linha 10 pela função $\mathbf{classify}_{E_j}(\mathbf{x})$, que refere-se à classificação realizada no elemento x pelo classificador j). Após o processamento, cada classificador C_i irá fornecer um vetor $v_i = [y_i^1, \dots, y_i^p]$, de p posições, onde cada elemento y_i^m refere-se à possibilidade do elemento X pertencer à classe m , de acordo com o Classificador i (este processo está descrito nas linhas 9–10 do algoritmo 2). Este processo será feito para classificador, formando uma matriz $n \times p$ (representada no algoritmo na linha 7) preenchida com os seguintes valores:

$$V = \begin{bmatrix} y_1^1 & y_1^2 & \dots \\ \vdots & \ddots & \\ y_n^1 & & y_n^p \end{bmatrix}$$

Na etapa de agregação, a AC Integral será aplicada utilizando pares funções de fusão de aridade 2 F_1 e F_2 e medidas *fuzzy* \mathbf{m}_1 e \mathbf{m}_2 para agregar os resultados referentes a cada classe que foi processada pelos classificadores. Em outras palavras, iremos obter, para

Algoritmo 3: Pseudo-código para cálculo da AC-Integral

```

1 Input:  $\mathbf{x} \leftarrow 0, x_1, \dots, x_n$  - É um vetor de tamanho  $n$ 
2  $x_s \leftarrow \text{sort}(\mathbf{x})$ 
3  $\mathbf{I} \leftarrow \text{original-position}(x_s)$ 
4  $y \leftarrow 0$ 
5 aggregate-ac-integral( $\mathbf{x}$ ):
6 for  $i = 1$  to  $n$  do
7    $y \leftarrow y + |F_1(x_s[i], m_1((I[i], \dots, I[n]))) - F_2(x_s[i - 1], m_2((I[i], \dots, I[n])))|$ 
8 end
9 return  $y$ 

```

cada classe w o valor $AC_{m_1, m_2}^{F_1 F_2}([y_1^w, \dots, y_n^w])$, o que na prática, é realizar a agregação das colunas da matriz V (processo que está descrito nas linhas 12 – 14 do Algoritmo 2).

Na linha 13, pode-se verificar a chamada à função **aggregate-ac-integral**, que está descrita no Algoritmo 3 para realizar o cálculo da AC-Integral. Nela, utilizamos as funções **sort**, e **original-position**. A função **sort** refere-se a uma função de ordenação crescente qualquer e a função **original-position** retorna à posição original dos elementos no vetor de entrada antes dele ser ordenado. Finalmente, após realizar a agregação para todas as classes, obtemos o seguinte vetor:

$$Y = \begin{pmatrix} AC_{m_1, m_2}^{F_1 F_2}([y_1^1, \dots, y_n^1]) \\ AC_{m_1, m_2}^{F_1 F_2}([y_1^2, \dots, y_n^2]) \\ \vdots \\ AC_{m_1, m_2}^{F_1 F_2}([y_1^w, \dots, y_n^w]) \end{pmatrix}$$

A última etapa do algoritmo (descrito na linha 15), o valor a ser retornado como a classe de X refere-se à posição com maior valor em Y .

5.2 Exemplo de Funcionamento da Arquitetura Proposta

Nesta seção, será mostrado um exemplo de funcionamento do algoritmo proposto, de forma a clarificar o seu funcionamento e mostrar na prática como se dá o processo de classificação utilizando o algoritmo proposto.

Seja X um elemento a ser classificado, então considere um comitê com 4 classificadores $\mathfrak{C} = \{C_1, C_2, C_3, C_4\}$ e um total de 4 possíveis classes $\{w_1, w_2, w_3, w_4\}$. Neste exemplo, iremos utilizar $F_1(x, y) = xy^2$ e $F_2(x, y) = x^2y$ e utilizaremos a medida *fuzzy* uniforme

para \mathbf{m}_1 e \mathbf{m}_2 .

O primeiro passo é realizar a classificação. Suponha que ao final da etapa de classificação, os classificadores retornaram os seguintes valores para classificação do elemento: $C_1(X) = [0.4, 0.2, 0.1, 0.3]$, $C_2(X) = [0.5, 0.3, 0.1, 0.1]$, $C_3(X) = [0.3, 0.3, 0.3, 0.1]$ e $C_4(X) = [0.1, 0.1, 0.3, 0.5]$. Os valores então são colocados na matriz V , que agrupa os resultados das classificações:

Tabela 4: Matriz V com o resultado das classificações

	w_1	w_2	w_3	w_4
C_1	0.4	0.2	0.1	0.3
C_2	0.5	0.3	0.1	0.1
C_3	0.3	0.3	0.3	0.1
C_4	0.1	0.1	0.3	0.5

O próximo passo é calcular o vetor Y , que consiste em agregar os valores para cada classe w_i . Inicialmente, calcularemos a agregação referente à classe $w_1 : [0.4, 0.5, 0.3, 0.1]$. Para realizar a agregação seguiremos o Algoritmo 3 com as etapas descritas a seguir:

- **Passo 1:** Ordena-se o vetor de entrada, obtendo um vetor ordenado $x_s = [0.1, 0.3, 0.4, 0.5]$
- **Passo 2:** Calcula-se o vetor com as posições originais dos valores do vetor ordenado, obtendo: $I = [4, 3, 1, 2]$
- **Passo 3:** Em seguida, dá-se início ao laço das linhas 4 – 7:
 - Para $i = 1$:

$$\begin{aligned}
 y &= y + |F_1(x_s[1], \mathbf{m}_U(\{4, 3, 2, 1\})) - F_2(x_s[0], \mathbf{m}_U(\{4, 3, 2, 1\}))| \\
 &= y + |F_1(0.1, 1) - F_2(0, 1)| \\
 &= y + |0.1 * 1^2 - 0^2 * 1| \\
 &= y + 0.1
 \end{aligned}$$

- Para $i = 2$:

$$\begin{aligned}
 y &= y + |F_1(x_s[2], \mathbf{m}_U(\{3, 2, 1\})) - F_2(x_s[1], \mathbf{m}_U(\{3, 2, 1\}))| \\
 &= y + |F_1(0.3, 0.75) - F_2(0.1, 0.75)| \\
 &= y + |0.3 * 0.75^2 - 0.1^2 * 0.75| \\
 &= y + 0.16125
 \end{aligned}$$

– Para $i = 3$:

$$\begin{aligned}
 y &= y + |F_1(x_s[3], \mathbf{m}_U(\{2, 1\})) - F_2(x_s[2], \mathbf{m}_U(\{2, 1\}))| \\
 &= y + |F_1(0.4, 0.5) - F_2(0.3, 0.5)| \\
 &= y + |0.4 * 0.5^2 - 0.3^2 * 0.5| \\
 &= y + 0.055
 \end{aligned}$$

– Para $i = 4$:

$$\begin{aligned}
 y &= y + |F_1(x_s[4], \mathbf{m}_U(\{1\})) - F_2(x_s[3], \mathbf{m}_U(\{1\}))| \\
 &= y + |F_1(0.5, 0.25) - F_2(0.4, 0.25)| \\
 &= y + |0.5 * 0.25^2 - 0.4^2 * 0.25| \\
 &= y + 0.00875
 \end{aligned}$$

O valor inicial de y é 0. Assim, ao final do laço, teremos que $Y[1] = 0 + 0.1 + 0.16125 + 0.055 + 0.00875 = 0.325$. Repetindo os passos descritos anteriormente para as outras classes e realizando os cálculos da mesma forma, obtemos os seguintes valores: $Y[2] = 0.26375$, $Y[3] = 0.2225$ e $Y[4] = 0.2275$. Finalmente, obtemos o seguinte vetor Y :

$$Y = \begin{pmatrix} 0.32500 \\ 0.26375 \\ 0.22250 \\ 0.22750 \end{pmatrix}$$

Como pode-se verificar, o maior valor no vetor Y está na posição 1 e refere-se à classe w_1 (0.32500). Desta forma, seria atribuído ao elemento X à classe w_1 .

5.3 Procedimento para Definição de Pesos

A utilização de pesos em sistemas de comitês classificadores é uma abordagem comum e muito utilizada, pois permite atribuir um peso maior para classificadores que têm melhor performance. Desta forma, iremos propor uma outra contribuição nesta tese, que será a proposta de um método para calcular os pesos automaticamente. A abordagem aqui definida é baseada na acurácia dos classificadores, definindo essencialmente uma etapa extra

de treinamento que ocorre antes do uso de um conjunto de dados com um determinado classificador.

Considere um conjunto de $\mathfrak{C} = \{C_1, \dots, C_n\}$ classificadores. Inicialmente, é necessário realizar uma amostragem estratificada do conjunto de dados a ser classificado. Isto é necessário particularmente em grandes bases de dados, onde o tempo de classificação pode ser muito demorado. Com esse subconjunto de dados é feita a etapa de treinamento dos classificadores e realizada a classificação com um conjunto de treinamento, que não necessariamente é o mesmo utilizado para treinar os classificadores.

Resultante desta etapa de classificação, calculamos a acurácia de cada classificador relativa ao conjunto de teste, obtendo um vetor de acurácias $v_{\%} = [x_1, \dots, x_n]$. O método proposto funciona atribuindo pesos com base no vetor de acurácias v , baseado em uma função monotônica g onde se $g(x) = x, g(0) = 0$ e $g(1) = 1$.

Algoritmo 4: Pseudo-código para o modelo proposto para definição de pesos

```

1 Entrada:  $\vec{X} = (x_1, \dots, x_n)$  - um vetor de  $n$  valores onde  $\forall x \in X, 0 \leq x \leq 1$ 
2 Saída:  $\vec{W} = (w_1, \dots, w_n)$  - um vetor com  $n$  pesos
3 calculate-weights ( $x$ ):
4  $u_1, \dots, u_n \leftarrow [0, \dots, 0]$  um vetor de  $L$  posições inicialmente vazio
5  $g \leftarrow$  uma função monotônica
6 for  $i = 1$  to  $n$  do
7    $u_i = g\left(\frac{x_i}{\sum_{j=1}^n x_j}\right)$ 
8 end
9 for  $i = 1$  to  $n$  do
10   $w_i = \frac{u_i}{\sum_{j=1}^n u_j}$ 
11 end
12 return  $\vec{W}$ 

```

O Algoritmo 4 mostra o pseudo-código para o cálculo de pesos proposto. O algoritmo consiste em inicialmente calcular o valor proporcional de cada elemento em relação ao vetor, utilizando a função g como um agente moderador que pode, por exemplo, ser utilizada para acentuar os valores maiores, ou serem utilizadas funções como a função sigmóide, que praticamente anularia classificadores de menor performance. Este processo está nas linhas 6–8 do Algoritmo 4. As linhas 9–10 são essencialmente uma normalização, trazendo os valores de volta para o intervalo $[0, 1]$.

Este algoritmo será utilizado em seções posteriores ao aplicar a medida *fuzzy* ponderada com o comitê de classificador proposto e a AC-Integral, pois permite a criação de

quantas medidas *fuzzy* forem necessárias, visto que a escolha de diferentes funções g cria medidas completamente distintas.

5.4 Configuração dos Experimentos

Nesta seção serão descritos os parâmetros utilizados para a realização dos testes para verificação da eficácia dos métodos propostos. Inicialmente serão mostradas as bases de dados utilizadas e em seguida detalhes sobre os testes que serão realizados e como foram implementados.

Com relação à validação dos dados, serão utilizados, como testes estatísticos os testes de Friedman e de Nemenyi. Conforme explicado no Capítulo 2, para melhor visualizar os resultados pareados obtidos pelo teste de Nemenyi, será utilizado o diagrama CD.

As métricas adotadas para validação dos resultados serão as também citadas no Capítulo 2: Acurácia, Precisão, Recall, Medida F-1 (Medida F_β com $\beta = 1$) e o Kappa de Cohen. Para uma análise mais concisa, os testes estatísticos serão aplicados somente sobre a Acurácia.

5.4.1 Bases de Dados

Para realização dos testes, foram selecionadas 23 bases do repositório de aprendizagem de máquina da Universidade da Califórnia Irvine (UCI Machine Learning Repository) (DUA; GRAFF, 2017). Nestas bases, foram realizados os seguintes pré-processamentos:

- **Remoção de valores faltosos:** Foram removidas das bases de dados elementos que contivessem valores em branco ou não informados. Desta forma, alguma das bases mostradas terão um número de instancias inferior ao existente na fonte.
- **Normalização:** Em todas as bases escolhidas, foi realizada a normalização dos valores para cada atributo.

A Tabela 5 descreve todos os parâmetros das bases de dados selecionadas para este trabalho. No Apêndice C está disponível a descrição detalhada de cada uma destas bases.

Tabela 5: Descrição das bases de dados utilizadas neste trabalho

Nome da Base	Total de Elementos	Número de Atributos	Número de Classes
monk's problems 1(M1)	432	6	2
multiple features - fourier (MF)	2000	76	10
statlog (ST)	2000	36	6
spam (SB)	4601	57	2
german-credit(GC)	1000	24	2
optical recognition (OP)	5620	64	10
cnae-9 (C9)	1080	858	9
splice (SP)	3190	287	3
heart statlog (HS)	270	13	2
winequality-red (WR)	1599	11	6
transfusion (TR)	747	4	2
bupa (BU)	345	6	2
haberman (HA)	306	3	2
balance scale (BS)	625	4	3
iris (IR)	150	4	3
breast cancer (BC)	683	9	2
glass (GL)	214	10	6
ionosphere(IO)	351	34	2
car evaluation (CE)	1728	21	4
nursery (NU)	12960	26	5
waveform (WA)	5000	21	3
vehicle (VE)	846	18	4
winequality-white (WW)	4898	11	7

5.4.2 Configurações de Parâmetro e do Comitê

Nos experimentos realizados no capítulo seguinte, serão utilizados comitês heterogêneos compostos por 4 classificadores: Naive Bayes (NB), Árvore de Decisão (AD), k-NN e Redes Neurais com o algoritmo Multilayer Perceptron (MLP). A escolha por estes classificadores especificamente se deu pela sua presença constante em muitos trabalhos encontrados na literatura. Em relação à implementação, todos os testes realizados nesta tese foram implementados em Python, utilizando a biblioteca *scikit-learn* (PEDREGOSA et al., 2011) e suas implementações dos métodos citados. Naturalmente, as implementações das AC e CQO-Integrais propostas não utilizaram os módulos destas bibliotecas, sendo implementadas pelo autor.

A proporção de classificadores no comitê será de 3:3:3:1 em relação aos classificadores MLP, AD, k-NN e NB, respectivamente. Especificamente para o teste com 5 comitês, será utilizada a proporção 2:1:1:1. Ao todo, serão testadas 5 configurações de comitê obedecendo as proporções previamente estabelecidas: 5,10,15,20,40.

5.5 Considerações Finais

Neste Capítulo, foram mostradas as principais contribuições metodológicas desta Tese, principalmente com a descrição do algoritmo e modelo aqui propostos. Em seguida, foram descritos os parâmetros e configurações dos experimentos que serão realizados no capítulo seguinte.

6 Resultados

Neste capítulo, serão mostrados os resultados obtidos com os métodos propostos. Inicialmente, serão realizados testes com a CQO-Integral com diferentes funções e medidas *fuzzy*. Em seguida, serão realizados testes com a AC-Integral, utilizando duas medidas *fuzzy* distintas na integral, e estes resultados serão comparados com outras abordagens. Finalmente, o capítulo será encerrado com uma discussão e análise sobre os resultados obtidos.

Todos os testes realizados utilizaram a técnica de validação cruzada (k -folds, em inglês), que consiste em dividir um conjunto de dados em k partições e utilizar, em cada rodada, uma delas para testar o modelo e as $k - 1$ restantes para treinamento. Esse processo se repete até que todas as partições tenham sido utilizadas para teste uma única vez. Aqui, adotamos $k = 5$.

6.1 Determinação do número de classificadores no comitê

Um dos parâmetros mais importantes ao se trabalhar com comitês de classificadores é a definição do número de classificadores dentro do comitê. Para isso, fizemos testes preliminares com as configurações descritas no Capítulo 5. Avaliaremos inicialmente a performance da AC-Integral utilizando a melhor configuração obtida (em experimentos realizados posteriormente) utilizando as 23 bases.

A Figura 6 mostra a média das Acurácias obtidas com as bases em relação ao número de classificadores no comitê. Como podemos verificar, a configuração que obteve melhor performance foi a com 10 comitês. Também se observa que ao se aumentar o número de classificadores, a performance diminui. Desta forma, nos próximos experimentos adotaremos o número de 10 classificadores no comitê, e em seções futuras, analisaremos a performance dos métodos propostos com relação as bases individualmente quando se

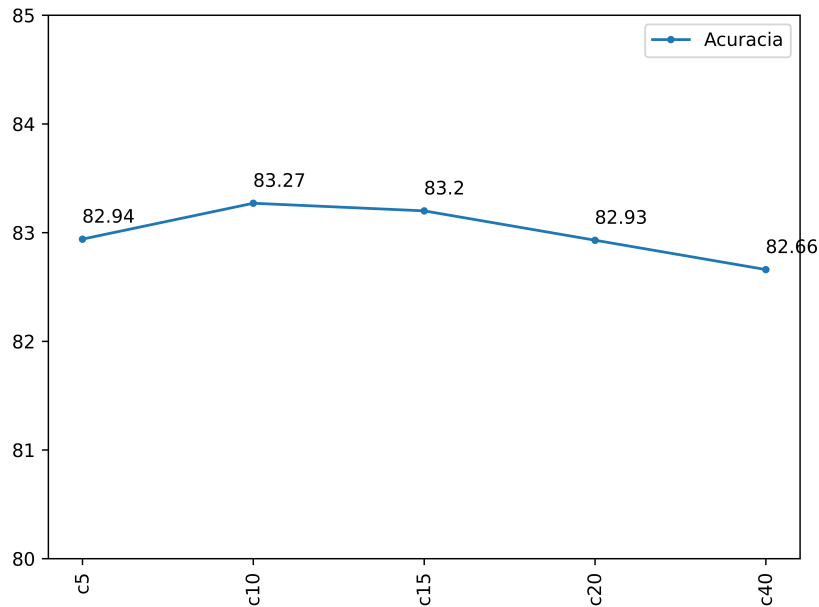


Figura 6: Comparação da performance do método ao variar a quantidade de classificadores presentes no comitê

muda o número de classificadores no comitê.

6.2 Experimentos com a CQO-Integral

Inicialmente, serão feitos testes com a CQO-Integral. Os resultados deste primeiro teste pode ser visto na Tabela 6. Nela, estão dispostos os resultados da aplicação da CQO-Integral utilizando alguns dos *quasi-overlaps* mostrados como exemplo na Tabela 2 no Capítulo 2, utilizando três medidas *fuzzy* diferentes: as medidas *fuzzy* λ , uniforme e relativa.

Analisando os resultados, pode-se verificar que o melhor resultado (82.11) foi obtido ao utilizar a função O_2 em conjunto com a medida *fuzzy* uniforme. Surpreendentemente, o melhor resultado foi obtido ao se utilizar a medida *fuzzy* uniforme, que é uma medida *fuzzy* relativamente simples. Também podemos verificar que o resultado obtido pelas funções O_σ e O_T foi idêntico, independente da medida *fuzzy* utilizada.

Para verificar se as diferenças entre os métodos foram estatisticamente significativas podemos primeiramente aplicar o teste de Friedman, que retorna um p -valor de $\approx 2.43e^{-06}$ que, obviamente, é suficientemente pequeno para descartar a hipótese nula da igualdade. Finalmente, podemos prosseguir para um teste pós-hoc, que neste caso é o teste de Nemenyi. Para representar visualmente a aplicação do teste e permitir uma análise compa-

Tabela 6: Resultados do experimento realizado com a CQO-Integral e a medida *fuzzy* uniforme

Medida Fuzzy Base	m_λ				m_U				m_R			
	O_σ	O_C	O_2	O_T	O_σ	O_C	O_2	O_T	O_σ	O_C	O_2	O_T
M1	94.89	94.89	95.36	94.89	92.33	92.57	92.33	92.33	92.57	92.57	92.57	92.57
MF	83.35	83.40	83.50	83.35	83.45	83.45	83.50	83.45	83.15	83.15	83.50	83.15
ST	84.30	84.40	84.15	84.30	84.40	84.45	84.50	84.40	84.30	84.25	84.30	84.30
SB	92.74	92.74	92.76	92.74	92.85	92.85	92.83	92.85	92.81	92.85	92.83	92.81
GC	75.00	74.90	74.60	75.00	75.50	74.90	75.50	75.50	74.80	74.90	75.20	74.80
OR	98.52	98.51	98.51	98.52	98.38	98.40	98.38	98.38	98.42	98.38	98.35	98.42
C9	95.09	95.00	94.91	95.09	94.81	94.54	94.44	94.81	94.54	94.44	94.54	94.54
SP	95.33	95.30	95.30	95.33	95.14	95.14	95.08	95.14	95.24	95.24	95.11	95.24
IR	95.33	95.33	95.33	95.33	96.00	95.33	96.00	96.00	95.33	95.33	95.33	95.33
HS	84.44	83.70	84.07	84.44	84.07	83.33	83.70	84.07	83.70	83.33	83.70	83.70
WR	54.53	54.72	53.47	54.53	55.35	55.41	55.66	55.35	54.53	54.91	55.35	54.53
TR	67.11	66.84	66.17	67.11	66.97	66.84	67.10	66.97	66.97	66.84	67.11	66.97
BU	73.33	73.33	73.33	73.33	73.62	72.17	73.62	73.62	72.17	72.17	73.04	72.17
HA	69.59	68.94	69.26	69.59	69.60	68.61	69.60	69.60	68.61	68.61	69.27	68.61
BS	80.64	80.64	79.52	80.64	78.88	78.08	79.20	78.88	77.12	77.60	78.88	77.12
BC	96.20	96.20	96.20	96.20	96.05	96.20	96.05	96.05	96.20	96.20	96.20	96.20
GL	93.42	92.96	92.95	93.42	93.43	92.49	93.44	93.43	93.42	92.96	92.97	93.42
IO	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90
CE	80.04	80.27	79.87	80.04	80.22	80.33	80.45	80.22	79.93	80.10	80.33	79.93
NU	71.03	71.15	71.02	71.03	71.10	71.07	71.08	71.10	71.08	71.11	71.03	71.08
WA	85.04	84.98	84.72	85.04	85.16	85.04	85.26	85.16	85.02	85.04	85.08	85.02
VE	77.90	77.78	77.78	77.90	78.01	78.02	78.37	78.01	77.31	77.42	78.13	77.31
WW	49.63	49.98	49.49	49.63	51.31	51.19	51.47	51.31	50.02	50.65	51.21	50.02
Média	82.10	82.04	81.88	82.10	82.07	81.80	82.11	82.07	81.66	81.69	81.95	81.66

rativa, será utilizado o diagrama CD, que está ilustrado na Figura 7.

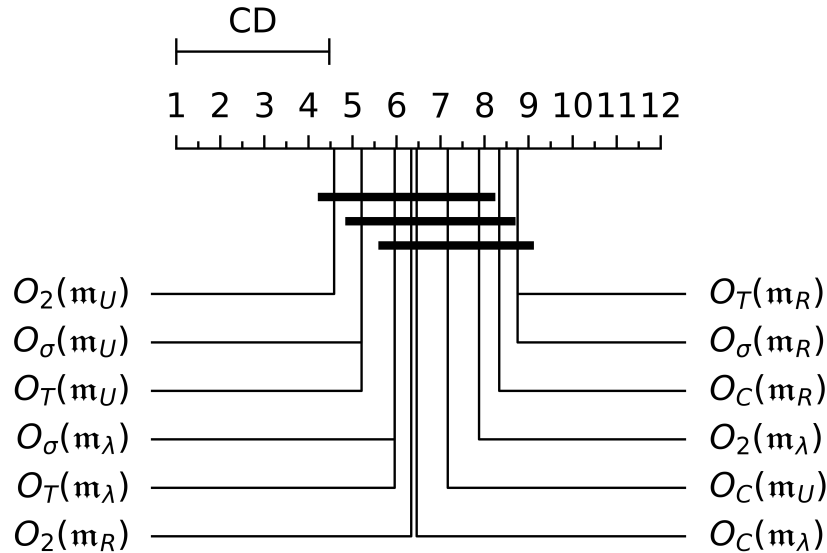


Figura 7: Diagrama CD para comparação dos métodos que utilizaram a integral CQO

Analisando os resultados, pode-se verificar que estatisticamente, a diferença entre muitos dos métodos não foi significativa. No entanto o método O_2 novamente se destacou, tendo o melhor ranque e tendo resultados superiores a alguns dos métodos que utilizaram

a medida fuzzy relativa (O_T , O_σ e O_C).

Para fins de comparação, foram implementados outros métodos e para eles, apresentados os mesmos dados e submetidos às mesmas condições dos métodos propostos. Os resultados estão dispostos na Tabela 7. Foram implementadas as seguintes abordagens: General Mixture (COSTA et al., 2018), Random Forests – RF (BREIMAN, 2001), AdaBoost – ADA (FREUND; SCHAPIRE, 1999), Bagging – BAG (BREIMAN, 1996), XGBoost – XGB (CHEN; GUESTRIN, 2016), META-DES – META (CRUZ et al., 2015) e RRC (WOLOSZYNSKI; KURZYNSKI, 2011). Além dos métodos externos, a tabela inclui a melhor combinação com a CQO-Integral para fins de comparação.

Tabela 7: Resultados da aplicação de outros métodos disponíveis na literatura

Base	$O_2(\mathbf{m}_U)$	GMM	RT	XGB	META	RRC	ADA	BAG
M1	92.33	87.23	69.45	74.81	91.87	93.03	61.57	51.38
MF	83.50	82.85	79.10	83.50	81.65	81.90	45.25	81.15
ST	84.50	83.65	83.50	83.35	79.95	80.05	63.40	83.25
SB	92.83	92.07	91.76	93.15	89.55	88.70	92.31	86.22
GC	75.50	74.20	74.70	75.80	68.60	68.20	76.30	71.20
OR	98.38	98.36	95.53	97.47	98.40	98.36	66.10	98.31
C9	94.44	93.89	90.56	90.83	93.80	94.17	37.41	84.17
SP	95.08	95.39	92.32	95.89	95.27	95.11	93.57	81.76
IR	96.00	96.00	96.67	96.00	96.00	95.33	94.67	96.00
HS	83.70	81.48	81.11	79.63	74.81	74.81	77.78	80.74
WR	55.66	54.85	53.22	54.60	49.41	49.53	50.84	51.78
TR	67.10	62.42	67.89	66.42	62.82	65.36	72.30	66.83
BU	73.62	70.43	68.41	69.86	66.38	65.51	69.28	63.19
HA	69.60	66.31	68.94	67.96	61.08	62.06	64.02	72.55
BS	79.20	74.08	64.16	69.76	70.88	72.48	69.28	75.84
BC	96.05	96.20	96.35	95.76	94.88	95.03	95.32	96.49
GL	93.44	89.67	85.05	85.38	89.67	90.60	73.32	83.19
IO	90.90	90.04	94.59	92.04	88.33	89.18	90.89	82.91
CE	80.45	78.65	76.74	78.37	77.96	78.30	73.95	72.87
NU	71.08	73.15	65.43	71.69	75.05	74.88	73.80	68.76
WA	85.26	84.14	81.28	84.90	77.56	76.98	84.02	81.72
VE	78.37	75.29	72.81	75.77	73.29	74.47	58.63	70.09
WW	51.47	49.02	48.86	51.14	42.79	42.94	39.61	48.29
Média	82.11	80.41	78.19	79.74	78.26	78.56	70.59	76.03

Comparando os resultados obtidos pela CQO-Integral com os outros métodos, verifica-se que os valores foram numericamente superiores. A média da acurácia do melhor resultado obtido com a CQO-Integral foi superior a todos os métodos externos. Além disso, a

CQO-Integra teve o melhor resultado em 12 das 23 bases.

Porém, resta saber se as diferenças obtidas são estatisticamente significativa. Através de uma análise estatística utilizando o teste de Friedman com os 6 métodos externos e o melhor resultado da CQO-Integra, obtém-se um p -valor menor que 0.001, que permite descartar a hipótese nula da igualdade e indica que existem diferenças estatisticamente significativa entre os dados. Aqui novamente será aplicado o diagrama CD para realizar a análise comparativa entre os métodos, que está ilustrado na Figura 8.

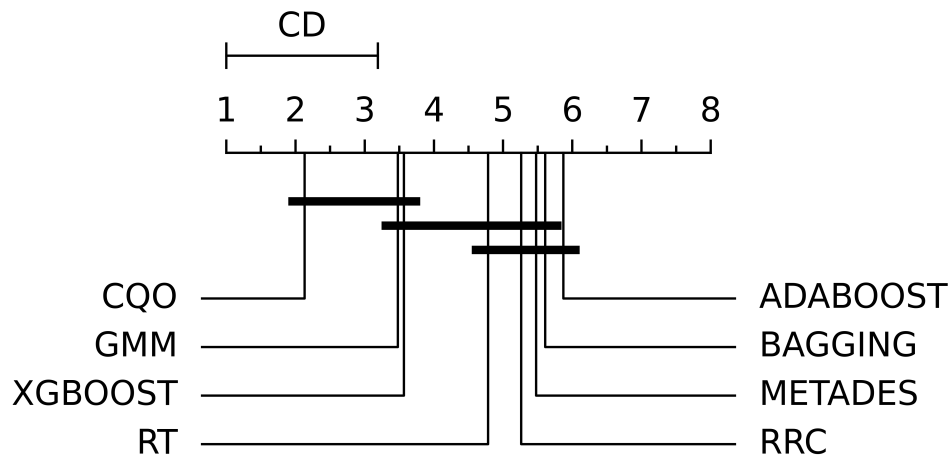


Figura 8: Diagrama CD para comparação entre os métodos externos e a melhor configuração da integral CQO

Analisando a Figura 8, podemos verificar que a diferença de desempenho entre a CQO Integral e os métodos externos foi estatisticamente significativa em todos os casos, exceto GMM e XGBoost. Estes resultados claramente indicam que a performance da CQO-Integra foi excelente ao classificar dados de diferentes bases. Finalmente, conforme discutido em seções anteriores, a acurácia nem sempre é método mais indicado para realizar as análises. Portanto, na Tabela 8 são mostrados resultados analisados sobre diferentes medidas descritas no Capítulo 2. Novamente, podemos verificar que a CQO-Integra teve o melhor resultado independentemente da métrica utilizada.

6.3 Experimentos com a AC-Integral

Nesta seção, serão apresentados os resultados obtidos com a AC-Integral. Como essa abordagem utiliza duas funções distintas, selecionamos algumas das funções utilizadas na Seção 6.2. Além disso, foram adicionados 4 *pseudo-quasi overlaps* que foram descritos na Tabela 3 do Capítulo 4.

Tabela 8: Resultados da aplicação de outros métodos disponíveis na literatura

	Acurácia	Medida F1	Precisão	Recall	Kappa
CQO	82.11	75.20	77.21	76.00	65.74
GMM	80.41	73.68	75.64	74.54	63.02
RT	78.19	70.81	73.49	71.56	59.63
XGB	79.74	72.99	75.12	73.59	62.53
META	78.26	72.02	73.20	72.97	60.04
RRC	78.56	72.27	73.38	73.20	60.41
ADA	70.59	60.86	64.40	63.08	50.60
BAG	76.03	67.81	69.69	68.38	55.04

No primeiro experimento, cujos resultados estão na Tabela 9, novamente foi utilizada a medida *fuzzy* uniforme tanto para \mathbf{m}_1 quanto para \mathbf{m}_2 . Testes com medidas *fuzzy* diferentes serão feitos adiante. Por enquanto, deseja-se verificar o resultado de usar funções distintas, de modo a verificar se houve um impacto positivo na classificação.

Tabela 9: Médias da acurácia dos resultados das classificações com a AC-Integral para cada par de combinações utilizando a medida uniforme

-	O_{PO}^1	O_{PO}^2	O_{PO}^3	O_{PO}^4	O_Q^1	O^1	C^1	O^2	O_Q^2
O_{PO}^1	82.00	82.34	82.48	81.83	81.73	82.15	81.87	81.84	80.97
O_{PO}^2	82.32	82.34	82.15	82.28	81.39	82.47	81.49	81.63	80.49
O_{PO}^3	82.35	82.48	82.33	82.27	81.37	82.47	81.39	81.50	80.16
O_{PO}^4	81.84	81.92	81.85	81.80	82.01	81.57	82.26	82.32	81.16
O_Q^1	81.81	81.28	81.28	81.78	82.11	81.69	81.47	81.35	82.01
O^1	82.05	82.57	82.53	81.35	81.80	82.07	82.07	82.08	77.68
C^1	81.90	81.38	81.38	82.01	82.08	81.82	82.17	82.06	82.14
O^2	81.88	81.41	81.44	82.04	82.02	81.83	82.23	82.07	82.15
O_Q^2	81.30	77.19	76.54	78.08	81.81	80.18	81.61	81.52	82.07

Na Tabela 9, pode-se verificar a média das classificações para cada par de funções utilizadas. Note que a AC-Integral não é simétrica, portanto, todas as possíveis combinações foram testadas. Para facilitar uma análise mais detalhada, visto que tem-se um total de 81 resultados, serão selecionadas algumas das combinações para serem analisadas frente a outras. Portanto, escolhemos as 5 melhores combinações ((O_{PO}^1, O_{PO}^3) , (O_{PO}^2, O^1) , (O_{PO}^3, O_{PO}^2) , (O^1, O_{PO}^2) e (O^1, O_{PO}^3)) para serem analisadas em mais detalhes. Os resultados estão dispostos na Tabela 10. As tabelas com todos os resultados estão no Apêndice A.

Na Tabela 10, podemos observar que os resultados foram numericamente superiores que os obtidos utilizando a CQO-Integral. Vale destacar que os melhores resultados que foram obtidos envolveram *pseudo-quasi overlaps*. Isso indica que a assimetria dessas funções pode ser um fator que contribua para a melhora do desempenho. Um outro ponto de

Tabela 10: Resultados do experimento realizado com a AC-Integral e a medida *fuzzy* uniforme

Base	(O_{PO}^1, O_{PO}^3)	(O_{PO}^2, O^1)	(O_{PO}^3, O_{PO}^2)	(O^1, O_{PO}^2)	(O^1, O_{PO}^3)
M1	90.25	89.32	92.57	92.10	91.40
MF	82.75	83.00	83.25	83.50	83.40
ST	84.05	84.15	84.45	84.30	84.30
SB	92.57	92.54	92.63	92.74	92.65
GC	75.20	75.00	75.00	75.40	75.00
OR	98.36	98.36	98.38	98.35	98.36
C9	95.09	94.91	95.09	95.00	94.91
SP	95.05	94.95	95.14	95.14	95.14
IR	96.00	96.00	96.00	96.00	96.00
HS	83.33	83.70	83.70	84.07	83.70
WR	57.85	58.29	56.54	56.22	56.47
TR	71.64	72.04	70.44	70.97	71.11
BU	72.46	72.75	73.62	73.91	73.91
HA	71.57	71.57	70.91	70.91	70.91
BS	80.80	80.32	80.16	80.16	80.16
BC	96.64	96.64	96.35	96.49	96.35
GL	93.44	93.44	93.44	93.91	93.44
IO	90.90	90.33	90.61	90.90	90.61
CE	80.79	80.68	80.85	80.56	80.85
NU	72.13	71.97	71.21	71.20	71.21
WA	86.02	86.04	85.66	85.64	85.60
VE	79.31	79.08	78.60	78.72	78.72
WW	52.94	53.14	52.49	50.96	52.67
Média	82.57	82.53	82.48	82.48	82.47

destaque é os melhores resultados também envolveram uma combinação de funções diferentes. O que indica que a AC-Integral também contribuiu para a melhora do desempenho da classificação.

Aplicando uma análise estatística com os dados obtidos na Tabela 10 e os obtidos pela CQO-Integral, obtemos um p -valor de 0.481, o que não é suficiente para descartar a hipótese nula da igualdade. Portanto, apesar de os resultados terem sido numericamente superiores, as diferenças não foram estatisticamente significativas.

No entanto, em geral, podemos considerar que a AC-Integral teve excelentes resultados, sendo numericamente superiores aos obtidos pela CQO-Integra, indicando que a utilização de duas funções distintas foi benéfica para a classificação. Em seguida, serão testadas as mesmas combinações porém utilizando a medida *fuzzy* ponderada. Os pesos serão determinados através do método apresentado no capítulo 4, cuja descrição está no Algoritmo 4.

Com relação às funções g utilizadas no método, foram selecionados duas funções $g_1(x) = x^{10}$ e $g_2(x) = x^8$. A escolha por funções com expoentes tão altos se deu com a intenção de acentuar as diferenças entre classificadores de menor performance dos que tiveram a melhor performance dentro do comitê. De maneira idêntica ao que foi feito para a Tabela 9, o comitê foi aplicado e os resultados dispostos na Tabela 11. As tabelas com todos os resultados estão no Apêndice B.

Tabela 11: Médias da acurácia dos resultados das classificações com a AC-Integral para cada par de combinações utilizando a medida *fuzzy* ponderada

-	O_{PO}^1	O_{PO}^2	O_{PO}^3	O_{PO}^4	O_Q^1	O^1	C^1	O^2	O_Q^2
O_{PO}^1	83.03	82.77	82.81	82.66	82.33	82.71	82.59	82.57	81.30
O_{PO}^2	83.09	83.24	83.26	83.27	81.58	83.17	81.62	81.81	81.72
O_{PO}^3	83.11	83.23	83.20	83.15	81.60	83.14	81.46	81.66	81.15
O_{PO}^4	82.95	82.79	82.61	82.94	82.22	82.92	82.44	82.46	81.54
O_Q^1	82.64	82.22	82.24	82.58	83.06	82.62	82.67	82.62	82.70
O^1	82.92	83.03	83.04	82.01	82.21	83.08	82.38	82.43	78.45
C^1	82.77	82.48	82.53	82.83	82.12	82.69	83.11	83.10	82.80
O^2	82.75	82.54	82.58	82.84	81.87	82.70	83.13	83.18	82.83
O_Q^2	82.13	78.80	78.49	78.69	82.53	81.05	82.25	82.18	83.08

Na Tabela 11, podemos observar que houve uma nova melhora através da utilização de pesos em relação a medida *fuzzy* uniforme. Por exemplo, para a combinação (O_{PO}^1, O_{PO}^1) a melhora foi de 1.03 na média da classificação (passou de 82.00 para 83.03). No entanto, uma única combinação teve uma pequena diminuição na taxa de acerto. No caso, esta redução correu na combinação (O^2, O_Q^1) , que passou de 82.02 para 81.87, uma piora de

0.15. No entanto, em geral, as combinações tiveram uma melhora sensível de desempenho. Na Tabela 12, pode-se verificar os ganhos das combinações.

Tabela 12: Ganho de desempenho das combinações após a utilização da medida *fuzzy* ponderada

-	O_{PO}^1	O_{PO}^2	O_{PO}^3	O_{PO}^4	O_Q^1	O^1	C^1	O^2	O_Q^2
O_{PO}^1	1.03	0.43	0.32	0.83	0.59	0.56	0.72	0.72	0.33
O_{PO}^2	0.77	0.90	1.11	0.98	0.19	0.70	0.13	0.18	1.23
O_{PO}^3	0.76	0.74	0.87	0.88	0.22	0.67	0.07	0.15	0.99
O_{PO}^4	1.11	0.87	0.76	1.14	0.22	1.35	0.18	0.15	0.39
O_Q^1	0.84	0.94	0.96	0.80	0.95	0.92	1.20	1.26	0.69
O^1	0.87	0.46	0.51	0.66	0.41	1.02	0.31	0.35	0.77
C^1	0.87	1.10	1.15	0.82	0.05	0.87	0.95	1.04	0.66
O^2	0.88	1.12	1.15	0.80	-0.15	0.87	0.90	1.11	0.68
O_Q^2	0.84	1.61	1.95	0.61	0.73	0.88	0.64	0.66	1.02

Claramente, podemos ver que a utilização da medida *fuzzy* ponderada melhorou, pelo menos numericamente, os resultados com relação à utilização da medida *fuzzy* uniforme. Em seguida, serão comparados as melhores configurações da AC-Integral utilizando pesos (O_{PO}^2, O_{PO}^4) com a melhor combinação utilizando a CQO-Integral e os métodos externos. Estes resultados podem ser visualizados na Tabela 13.

Comparando os resultados obtidos, verifica-se que a AC-Integral (ACI) obteve os melhores resultados em 12 das 23 bases, perdendo em algumas instâncias para a CQO-Integral. Se consideramos apenas os algoritmos propostos nesta tese, foram obtidos resultados superiores em 16 das 23 bases. Prosseguindo para uma análise estatística, ao aplicar o teste de Friedman obtemos um p -valor menor que 0.01, permitindo descartar a hipótese nula da igualdade. A aplicação e visualização do teste de Nemenyi será novamente ilustrada por meio do diagrama CD, mostrado na Figura 9.

Analisando o Diagrama CD, verificamos que a AC-Integral teve uma colocação superior a CQO-Integral, mas as diferenças não foram estatisticamente significativas. Além, disso, a AC-Integral teve diferenças estatisticamente significativas com relação aos métodos Adaboost, Bagging, META-DES, RRC e RT. Com relação aos outros métodos as diferenças não foram estatisticamente significativas, embora tenha havido um sensível aumento na média da classificação. Para verificar os resultados novamente por meio de outras métricas, utilizamos novamente as métricas definidas anteriormente para comparar os métodos. Podemos visualizar essa comparação na Tabela 14, onde podemos verificar que a AC-Integral obteve os melhores resultados em todas as métricas.

Em geral, podemos considerar que a AC-Integral obteve bons resultados quando com-

Tabela 13: Resultados da aplicação de outros métodos disponíveis na literatura

Base	ACI	CQO	GMM	RT	XGB	META	RRC	ADA	BAG
M1	89.10	92.33	87.23	69.45	74.81	91.87	93.03	61.57	51.38
MF	83.69	83.50	82.85	79.10	83.50	81.65	81.90	45.25	81.15
ST	84.00	84.50	83.65	83.50	83.35	79.95	80.05	63.40	83.25
SB	92.89	92.83	92.07	91.76	93.15	89.55	88.70	92.31	86.22
GC	76.10	75.50	74.20	74.70	75.80	68.60	68.20	76.30	71.20
OR	98.68	98.38	98.36	95.53	97.47	98.40	98.36	66.10	98.31
C9	95.46	94.44	93.89	90.56	90.83	93.80	94.17	37.41	84.17
SP	95.36	95.08	95.39	92.32	95.89	95.27	95.11	93.57	81.76
IR	96.00	96.00	96.00	96.67	96.00	96.00	95.33	94.67	96.00
HS	83.70	83.70	81.48	81.11	79.63	74.81	74.81	77.78	80.74
WR	55.66	55.66	54.85	53.22	54.60	49.41	49.53	50.84	51.78
TR	73.38	67.10	62.42	67.89	66.42	62.82	65.36	72.30	66.83
BU	73.91	73.62	70.43	68.41	69.86	66.38	65.51	69.28	63.19
HA	75.82	69.60	66.31	68.94	67.96	61.08	62.06	64.02	72.55
BS	93.12	79.20	74.08	64.16	69.76	70.88	72.48	69.28	75.84
BC	96.64	96.05	96.20	96.35	95.76	94.88	95.03	95.32	96.49
GL	91.52	93.44	89.67	85.05	85.38	89.67	90.60	73.32	83.19
IO	90.33	90.90	90.04	94.59	92.04	88.33	89.18	90.89	82.91
CE	80.22	80.45	78.65	76.74	78.37	77.96	78.30	73.95	72.87
NU	71.47	71.08	73.15	65.43	71.69	75.05	74.88	73.80	68.76
WA	86.76	85.26	84.14	81.28	84.90	77.56	76.98	84.02	81.72
VE	81.56	78.37	75.29	72.81	75.77	73.29	74.47	58.63	70.09
WW	49.82	51.47	49.02	48.86	51.14	42.79	42.94	39.61	48.29
Média	83.27	82.11	80.41	78.19	79.74	78.26	78.56	70.59	76.03

Tabela 14: Diferença entre o desempenho da melhor configuração da AC-Integral contra outros métodos

	Acurácia	Medida F1	Precisão	Recall	Kappa
ACI	83.27	76.95	79.96	77.57	67.91
CQO	82.11	75.20	77.21	76.00	65.74
GMM	80.41	73.68	75.64	74.54	63.02
RT	78.19	70.81	73.49	71.56	59.63
XGB	79.74	72.99	75.12	73.59	62.53
META	78.26	72.02	73.20	72.97	60.04
RRC	78.56	72.27	73.38	73.20	60.41
ADA	70.59	60.86	64.40	63.08	50.60
BAG	76.03	67.81	69.69	68.38	55.04

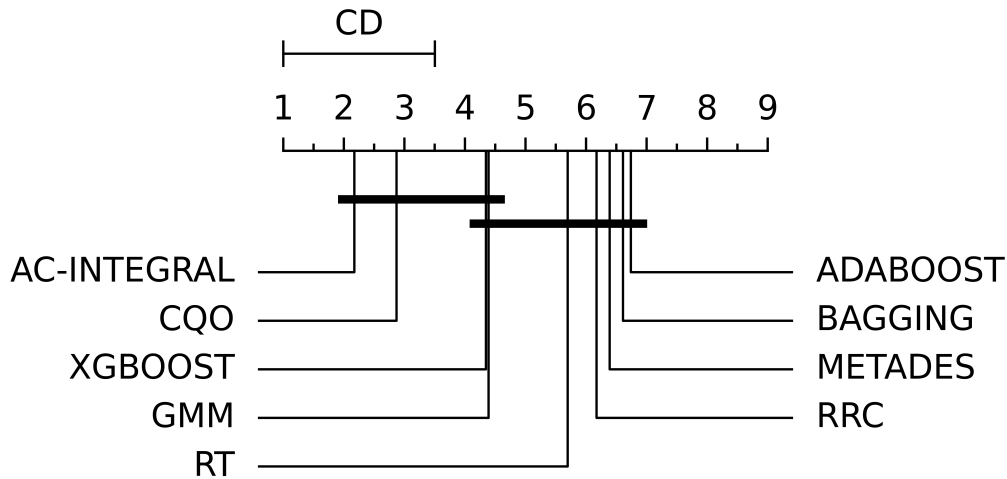


Figura 9: Diagrama CD para comparação entre os métodos externos e a melhor configuração da CQO-Integral e AC-Integral

parado com outros métodos. Embora as diferenças entre a AC-Integral e o Método XGBoost e GMM não tenham sido estatisticamente significativas, o método XGBoost é um método que vem obtendo excelentes resultados e tem sido utilizado como referência para muitos trabalhos, e o método GMM foi um método recente proposto que obteve resultados similares ao XGBoost. Apesar disso, os resultados foram numericamente superiores em múltiplas análises.

6.4 Discussão dos resultados

De forma geral, os resultados obtidos pelos métodos propostos foram superiores aos outros métodos. Isto foi discutido e apresentado através de diversas comparações e métricas em seções anteriores. Nesta seção, serão exploradas, de forma empírica, as condições e fatores que parecem ter contribuído para uma melhor performance da AC-Integral. Inicialmente, na Tabela 15, podemos verificar a diferença de desempenho (utilizando a métrica da Acurácia) entre a AC-Integral (utilizando a combinação (O_{PO}^2, O_{PO}^4) e a medida *fuzzy* ponderada). As duas últimas colunas representam a média dos ganhos para os métodos do estado da arte – EdA (GMM,RT,XGB,META e RRC) e para os métodos clássicos – MC (ADA e BAG), respectivamente

Analisando a Tabela 15, vemos que as principais bases onde houve ganho de performance em relação aos classificadores do estado da arte (EdA) foram: BS, HA, TR, VE e M1. Analisando a Tabela 5, o que se encontra de comum entre as bases selecionadas é o número reduzido de classes (todas com um número menor que 4), o que indica que a

Tabela 15: Resultados da aplicação de outros métodos disponíveis na literatura

Dataset	GMM	RT	XGB	META	RRC	ADA	BAG	Média (EdA)	Média (MC)
M1	1.87	19.65	14.29	-2.77	-3.93	27.53	37.72	5.82	32.63
MF	0.75	4.50	0.10	1.95	1.70	38.35	2.45	1.80	20.40
ST	0.35	0.50	0.65	4.05	3.95	20.60	0.75	1.90	10.68
SB	0.82	1.13	-0.26	3.34	4.19	0.58	6.67	1.84	3.63
GC	1.90	1.40	0.30	7.50	7.90	-0.20	4.90	3.80	2.35
OR	0.32	3.15	1.21	0.28	0.32	32.58	0.37	1.06	16.48
C9	1.57	4.90	4.63	1.66	1.29	58.05	11.29	2.81	34.67
SP	-0.03	3.04	-0.53	0.09	0.25	1.79	13.60	0.56	7.70
IR	0.00	-0.67	0.00	0.00	0.67	1.33	0.00	0.00	0.67
HS	2.22	2.59	4.07	8.89	8.89	5.92	2.96	5.33	4.44
WR	0.81	2.44	1.06	6.25	6.13	4.82	3.88	3.34	4.35
TR	10.96	5.49	6.96	10.56	8.02	1.08	6.55	8.40	3.82
BU	3.48	5.50	4.05	7.53	8.40	4.63	10.72	5.79	7.68
HA	9.51	6.88	7.86	14.74	13.76	11.80	3.27	10.55	7.54
BS	19.04	28.96	23.36	22.24	20.64	23.84	17.28	22.85	20.56
BC	0.44	0.29	0.88	1.76	1.61	1.32	0.15	1.00	0.74
GL	1.85	6.47	6.14	1.85	0.92	18.20	8.33	3.45	13.27
IO	0.29	-4.26	-1.71	2.00	1.15	-0.56	7.42	-0.51	3.43
CE	1.57	3.48	1.85	2.26	1.92	6.27	7.35	2.22	6.81
NU	-1.68	6.04	-0.22	-3.58	-3.41	-2.33	2.71	-0.57	0.19
WA	2.62	5.48	1.86	9.20	9.78	2.74	5.04	5.79	3.89
VE	6.27	8.75	5.79	8.27	7.09	22.93	11.47	7.23	17.20
WW	0.80	0.96	-1.32	7.03	6.88	10.21	1.53	2.87	5.87

AC-Integral possui um desempenho melhor quando o espaço de possíveis decisões é reduzido. Vale destacar, porém, que isso não significa que a performance do comitê é inferior as outras abordagens em outras situações, conforme pode ser visto na tabela referenciada.

Apenas em duas das 23 bases o desempenho da AC-Integral foi inferior: IO e NU, ainda que a diferença tenha sido de aproximadamente 0.5%. À primeira vista, não há nenhum parâmetro em comum entre essas duas bases, além de um número razoável de atributos (34 e 26, respectivamente). Porém, existem bases com um número maior de atributos, como a C9, em que houve um aumento do desempenho. Em geral, a única indicação que pode-se inferir a partir dos experimentos conduzidos é que um número menor de classes parece favorecer a performance da AC-Integral.

Com relação ao tamanho do comitê, podemos verificar o impacto do aumento no número de classificadores através da Figura 10. Nela, podemos ver, por base de dados, de como a mudança deste parâmetro afetou a acurácia, utilizando configurações de 5, 10, 15, 20 e 40 classificadores. Como método, escolhemos a melhor configuração da AC-Integral utilizando a medida *fuzzy* uniforme. O que se observa através de uma análise inicial é que houve uma tendência de queda em algumas bases conforme o tamanho do comitê

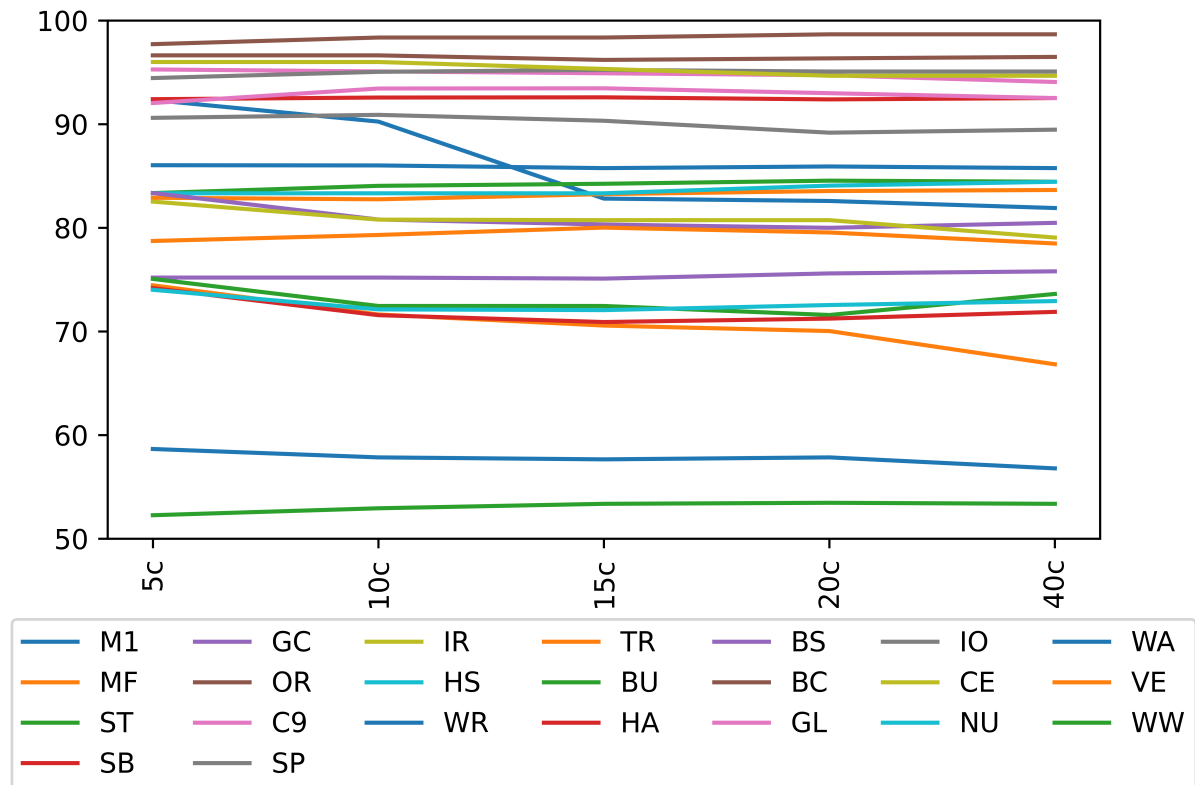


Figura 10: Comparação da acurácia do comitê pela quantidade de classificadores presentes no comitê

aumenta.

A tendência de queda é ilustrada principalmente pelas bases WA (azul) e VE (laranja) que observaram uma queda acentuada. Desta forma, em geral, concluímos a performance dos comitês não parece ter se beneficiado de um aumento do número de classificadores, portanto, isso indica que a abordagem proposta nesse trabalho consegue trabalhar e obter bons resultados com um número reduzido de classificadores no comitê.

6.5 Considerações Finais

Neste capítulo, apresentamos os resultados desta tese. Em geral, os resultados foram satisfatórios. Inicialmente, foi testada a CQO-Integral, que obteve bons resultados em relação aos métodos externos. Em seguida, foram apresentados os resultados para a AC-Integral utilizando a medida *fuzzy* uniforme e medida *fuzzy* ponderada. Todos os resultados apresentados foram seguidos de análises estatísticas. Finalmente, foi apresentada a discussão dos resultados.

7 Considerações Finais

Neste capítulo serão apontadas algumas considerações para conclusão desta tese, seguidas de possíveis ramificações e trabalhos futuros que podem ser realizados a partir dos resultados obtidos.

7.1 Considerações

Nesta tese, foi trabalhado o conceito de Comitês de Classificadores, onde o foco foi propor uma nova abordagem focada na camada de combinação. Para tal, foram propostas generalizações das integrais de Choquet para serem utilizadas como método de combinação nestes sistemas. Inicialmente, propomos novas funções de agregação, chamadas pseudo-quasi *overlaps* e *pseudo overlaps*, que são extensões das funções de *overlap* em que a necessidade da comutatividade foi removida (no caso dos pseudo *overlaps*) ou tanto a comutatividade quanto a continuidade (no caso dos *pseudo-quasi overlaps*). Do ponto de vista teórico, ainda foi proposto um novo tipo de monotonicidade direcional.

Baseado na ideia de generalizações das Integrais de Choquet, foram propostas três novas generalizações:

- **CQO-Integral:** Esta generalização consistiu na substituição de ambos os produtos presentes na Integral de Choquet expandida (Equação 2.22) por *quasi-overlaps*. A incorporação dos *quasi-overlaps* permitiu a utilização de funções que não necessariamente são contínuas no contexto da Integral de Choquet, além de permitir a incorporação dos *overlaps*, que já possuem histórico de aplicação em aplicações de aprendizagem de máquina (JURIO et al., 2013).
- **Pseudo-Integral de Choquet:** Esta generalização seguiu a mesma ideia da generalização anterior. No entanto, neste caso, o produto foi substituído por *pseudo overlaps*, que permitiram a integração de funções que não necessariamente são comutativas na Integral de Choquet.

- **AC-Integral:** Finalmente, a AC-Integral, inspirada pelas Integrais CF_1F_2 (LUCCA et al., 2019), substitui o produto na Integral de Choquet expandida por duas funções de agregação F_1 e F_2 . Neste caso, a AC-Integral permite a incorporação dos pseudo-quasi *overlaps*, que foram definidos anteriormente e que são, de forma simplificada, *overlaps* que não necessariamente são comutativos ou contínuos, o que permite a incorporação das características e benefícios citados nas ultimas duas funções. Além disso, essa generalização tratou alguns dos problemas existentes nas Integrais CF_1F_2 e em outras generalizações, como o uso de medidas *fuzzy* simétricas e a necessidade de dominância de F_1 sobre F_2 .

Para verificar a acurácia das generalizações propostas no contexto de comitês de classificadores, elas foram integradas na camada de combinação do comitê, servindo para agregar as opiniões relativas às classes dos elementos. Nos testes realizados, a CQO-Integral teve um bom desempenho, tendo tido uma melhor performance com relação a outros métodos conhecidos da literatura, a maioria deles com diferenças estatisticamente significativas.

Com relação a AC-Integral, os resultados obtidos também foram interessantes. Com aos métodos externos com os quais houve comparação, a AC-Integral teve desempenho numericamente superior, com diferenças estatisticamente significativas com relação a todos, exceto o XGBoost e o GMM. No entanto, como pode-se verificar na Figura 9, as diferenças estão no limite da distância crítica e portanto, caso tivesse sido adotado um nível de significância de 90%, as diferenças potencialmente poderiam ser consideradas como estatisticamente significativas.

Com relação aos testes com as funções de agregação propostas, em seções anteriores foi mostrado que os melhores resultados obtidos na AC-Integral foram obtidos ao se utilizar *pseudo-quasi overlaps*. Em especial, as funções selecionadas não eram nem contínuas nem comutativas. Uma possível explicação dos benefícios da remoção da necessidade da comutatividade é que na integral de Choquet substitui um produto, e os operandos tem diferentes origens, e possivelmente não tem a mesma importância. Portanto, a remoção da comutatividade permite tratar estes elementos de forma diferente.

Em geral, os resultados obtidos nesta tese foram satisfatórios. Embora a CQO-Integral tenha tido um desempenho inferior à AC-Integral, ela ainda é uma alternativa competitiva, tendo ganhado de métodos conhecidos na literatura. Em relação a AC-Integral, pode-se considerar que ela é uma abordagem muito promissora, que permite uma infinidade de otimizações e desdobramentos que podem ser explorados em trabalhos futuros. Além disso,

seu desempenho foi superior a maioria dos métodos com qual foi comparada.

7.2 Trabalhos Futuros

Aqui, fizemos estudos iniciais selecionando algumas funções escolhidas arbitrariamente. No entanto, em trabalhos futuros, deseja-se realizar um estudo para verificar quais funções ou famílias de funções maximizam a AC-Integral. Além disso, podem realizados testes com diferentes medidas *fuzzy* para verificar qual melhor se adequa a este novo método de combinação. Finalmente, deseja-se também fazer um estudo mais aprofundado sobre as relações entre as características das bases de dados utilizada e a performance da integral.

Além das ramificações citadas anteriormente, deseja-se analisar também o desempenho da pseudo-integral de Choquet como método de combinação em comitês de classificadores e comparar a performance generalizações das integrais de Choquet propostas com outras generalizações disponíveis na literatura sob diversos pontos de vista. Também pretende-se aprofundar o estudo teórico das generalizações das integrais de Choquet.

7.3 Produções Bibliográficas

Foi publicado um artigo e foram produzidos dois que estão em processo de revisão no momento da escrita desta tese:

- *Residuated implications derived from quasi-overlap functions on lattices* (PAIVA et al., 2021a) embora não faça parte explicitamente desta tese, serviu como fundação para o desenvolvimento de outras generalizações de funções de overlap.
- *Constructing Multi-layer Classifier Ensembles using the Choquet Integral based on Overlap and Quasi-Overlap Functions* (BATISTA; BEDREGAL; MORAES, 2022) consiste na apresentação da Integral CQO e comparação com outros métodos existentes na literatura.
- *Applying absolute Choquet integrals to classifier ensembles with a new class of aggregation functions* consiste na proposição da AC-Integral e Pseudo-Quasi Overlaps, também realizando testes com outros métodos. (BATISTA et al., 2022)

Referências

- ABDEL-NASSER, M.; MAHMOUD, K.; LEHTONEN, M. Reliable solar irradiance forecasting approach based on Choquet integral and deep LSTMs. *IEEE Transactions on Industrial Informatics*, v. 17, n. 3, p. 1873–1881, 2020.
- AGGARWAL, C. C. et al. *Neural networks and deep learning*. [S.l.]: Springer, 2018.
- ALSINA, C.; SCHWEIZER, B.; FRANK, M. J. *Associative functions: triangular norms and copulas*. [S.l.]: World Scientific, 2006.
- ALZUBI, O. A. et al. Consensus-based combining method for classifier ensembles. *Int. Arab J. Inf. Technol.*, v. 15, n. 1, p. 76–86, 2018.
- ATALLAH, R.; AL-MOUSA, A. Heart disease detection using machine learning majority voting ensemble method. In: IEEE. *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*. [S.l.], 2019. p. 1–6.
- BARZEGAR, R. et al. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stochastic Environmental Research and Risk Assessment*, v. 32, n. 3, p. 799–813, 2018.
- BATISTA, T.; BEDREGAL, B.; MORAES, R. Constructing multi-layer classifier ensembles using the Choquet integral based on overlap and quasi-overlap functions. *Neurocomputing*, p. To appear, 2022.
- BATISTA, T. et al. Applying absolute Choquet integrals to classifier ensembles with a new class of aggregation functions. *Fuzzy Sets and Systems*, p. To appear, 2022.
- BATISTA, T. V. et al. Farmyo: a serious game for hand and wrist rehabilitation using a low-cost electromyography device. *International Journal of Serious Games*, v. 6, n. 2, p. 3–19, 2019.
- BEDREGAL, B. et al. Generalized interval-valued OWA operators with interval weights derived from interval-valued overlap functions. *International Journal of Approximate Reasoning*, v. 90, p. 1–16, 2017.
- BELIAKOV, G.; JAMES, S.; WU, J. *Discrete fuzzy measures*. [S.l.]: Springer, 2020.
- BELIAKOV, G. et al. *Aggregation functions: A guide for practitioners*. [S.l.]: Springer, 2007.
- BREIMAN, L. Bagging predictors. *Machine learning*, v. 24, n. 2, p. 123–140, 1996.
- BREIMAN, L. Random forests. *Machine learning*, v. 45, n. 1, p. 5–32, 2001.

- BUSTINCE, H. et al. Ordered directionally monotone functions: Justification and application. *IEEE Transactions on Fuzzy Systems*, v. 26, n. 4, p. 2237–2250, 2018.
- BUSTINCE, H. et al. Overlap functions. *Nonlinear Analysis: Theory, Methods & Applications*, v. 72, n. 3-4, p. 1488–1499, 2010.
- BUSTINCE, H. et al. Overlap functions. *Nonlinear Analysis: Theory, Methods & Applications*, v. 72, n. 3, p. 1488–1499, 2010. ISSN 0362-546X.
- BUSTINCE, H. et al. d-Choquet integrals: Choquet integrals based on dissimilarities. *Fuzzy Sets and Systems*, v. 414, p. 1–27, 2021.
- BUSTINCE, H. et al. Pre-aggregation functions: definition, properties and construction methods. In: IEEE. *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. [S.l.], 2016. p. 294–300.
- CAO, Y. Aggregating multiple classification results using Choquet integral for financial distress early warning. *Expert Systems with Applications*, v. 39, n. 2, p. 1830–1836, 2012.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The Matthews correlation coefficient (MCC) is more informative than Cohen’s Kappa and brier score in binary classification assessment. *IEEE Access*, v. 9, p. 78368–78381, 2021.
- CHOQUET, G. Theory of capacities. In: *Annales de l’institut Fourier*. [S.l.: s.n.], 1954. v. 5, p. 131–295.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, v. 20, n. 1, p. 37–46, 1960.
- CONOVER, W. J. *Practical nonparametric statistics*. [S.l.]: John Wiley & sons, 1999.
- CORDEIRO, A. M. et al. Revisão sistemática: uma revisão narrativa. *Revista do Colégio Brasileiro de Cirurgiões*, v. 34, p. 428–431, 2007.
- COSTA, V. S. et al. Combining multiple algorithms in classifier ensembles using generalized mixture functions. *Neurocomputing*, v. 313, p. 402–414, 2018.
- CRUZ, R. M. et al. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern Recognition*, v. 48, n. 5, p. 1925–1935, 2015.
- CUNNINGHAM, P.; CORD, M.; DELANY, S. J. Supervised learning. In: *Machine learning techniques for multimedia*. [S.l.]: Springer, 2008. p. 21–49.
- CUNNINGHAM, P.; DELANY, S. k-nearest neighbour classifiers. *Mult Classif Syst*, v. 54, 04 2007.
- DAI, S.; CHENG, W. Noncommutative symmetric difference operators for fuzzy logic. *Journal of Intelligent & Fuzzy Systems*, v. 37, n. 6, p. 8005–8013, 2019.

- DANIEL, W. W.; CROSS, C. L. *Biostatistics: a foundation for analysis in the health sciences*. [S.l.]: Wiley, 2018.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.
- DEY, S. et al. Choquet fuzzy integral-based classifier ensemble technique for COVID-19 detection. *Computers in Biology and Medicine*, p. 104585, 2021.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *International Workshop on Multiple Classifier Systems*. [S.l.], 2000. p. 1–15.
- DIMURO, G. P. et al. The state-of-art of the generalizations of the Choquet integral: from aggregation and pre-aggregation to ordered directionally monotone functions. *Information Fusion*, v. 57, p. 27–43, 2020.
- DIMURO, G. P. et al. Generalized CF1F2-integrals: from Choquet-like aggregation to ordered directionally monotone functions. *Fuzzy Sets and Systems*, v. 378, p. 44–67, 2020.
- DING, S. et al. Pruning the ensemble of ann based on decision tree induction. *Neural Processing Letters*, v. 48, n. 1, p. 53–70, 2018.
- DOGAN, A.; BIRANT, D. A weighted majority voting ensemble approach for classification. In: IEEE. *2019 4th International Conference on Computer Science and Engineering (UBMK)*. [S.l.], 2019. p. 1–6.
- DUA, D.; GRAFF, C. *UCI Machine Learning Repository*. 2017. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- ELKANO, M. et al. Fuzzy rule-based classification systems for multi-class problems using binary decomposition strategies: on the influence of n-dimensional overlap functions in the fuzzy reasoning method. *Information Sciences*, v. 332, p. 94–114, 2016.
- FLONDOR, P.; GEORGESCU, G.; IORGULESCU, A. Pseudo-t-norms and pseudo-BL algebras. *Soft computing*, v. 5, n. 5, p. 355–371, 2001.
- FREUND, Y.; SCHAPIRE, R. E. A short introduction to boosting. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 1999. p. 1401–1406.
- GRABISCH, M. et al. Aggregation functions: construction methods, conjunctive, disjunctive and mixed classes. *Information Sciences*, v. 181, n. 1, p. 23–43, 2011.
- GREEN, B. N.; JOHNSON, C. D.; ADAMS, A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine*, v. 5, n. 3, p. 101–117, 2006.
- HAYKIN, S. A comprehensive foundation. *Neural Networks*, v. 2, n. 2004, p. 41, 2004.
- HONG, D. et al. Graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 59, n. 7, p. 5966–5978, 2021.
- HÜHN, J.; HÜLLERMEIER, E. FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, v. 19, n. 3, p. 293–319, 2009.

- JURIO, A. et al. Some properties of overlap and grouping functions and their application to image thresholding. *Fuzzy Sets and Systems*, v. 229, p. 69–90, 2013.
- KADKHODAEI, H. R.; MOGHADAM, A. M. E.; DEHGHAN, M. Hboost: A heterogeneous ensemble classifier based on the boosting method and entropy measurement. *Expert Systems with Applications*, v. 157, p. 113482, 2020.
- KARCZMAREK, P.; KIERSZTYN, A.; PEDRYCZ, W. Generalized Choquet integral for face recognition. *International Journal of Fuzzy Systems*, v. 20, n. 3, p. 1047–1055, 2018.
- KODOVSKY, J.; FRIDRICH, J.; HOLUB, V. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, v. 7, n. 2, p. 432–444, 2011.
- KRAWCZYK, B.; WOŹNIAK, M. Untrained weighted classifier combination with embedded ensemble pruning. *Neurocomputing*, v. 196, p. 14–22, 2016. ISSN 0925-2312.
- KRAWCZYK, B.; WOZNIAK, M. Leveraging ensemble pruning for imbalanced data classification. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.: s.n.], 2018. p. 439–444.
- KUMAR, G.; ROY, S. Development of hybrid boosting technique for bankruptcy prediction. In: IEEE. *2016 International Conference on Information Technology (ICIT)*. [S.l.], 2016. p. 248–253.
- KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, v. 51, n. 2, p. 181–207, 2003.
- LI, X.; WANG, F.; CHEN, X. Support vector machine ensemble based on Choquet integral for financial distress prediction. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 29, n. 04, 2015.
- LIN, X. et al. Performance analysis of pattern classifier combination by plurality voting. *Pattern Recognition Letters*, v. 24, n. 12, p. 1959–1969, 2003.
- LIU, H.; DU, Y.; WU, Z. Aem: Attentional ensemble model for personalized classifier weight learning. *Pattern Recognition*, v. 96, 2019.
- LIU, Z. et al. Classifier fusion with contextual reliability evaluation. *IEEE Transactions on Cybernetics*, v. 48, n. 5, p. 1605–1618, 2017.
- LLAMAZARES, B. Construction of Choquet integrals through unimodal weighting vectors. *International Journal of Intelligent Systems*, v. 33, n. 4, p. 771–790, 2018.
- LUCCA, G. et al. Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-Integrals named $C_{F_1 F_2}$ -Integrals. *IEEE Transactions on Fuzzy Systems*, v. 27, n. 1, p. 124–134, 2019.
- LUCCA, G. et al. Preaggregation functions: Construction and an application. *IEEE Transactions on Fuzzy Systems*, v. 24, n. 2, p. 260–272, 2016.

- LUCCA, G. et al. 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, 2017.
- LUO, C. A comparison analysis for credit scoring using bagging ensembles. *Expert Systems*, p. e12297, 2018.
- MAKKAR, T. et al. Analogizing time complexity of KNN and CNN in recognizing handwritten digits. In: IEEE. *2017 Fourth International Conference on Image Information Processing (ICIIP)*. [S.l.], 2017. p. 1–6.
- MEDSKER, L. R. *Hybrid neural network and expert systems*. [S.l.]: Springer Science & Business Media, 2012.
- MEZZOMO, I. et al. On the dominance relation between ordinal sums of quasi-overlap functions. In: IEEE. *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. [S.l.], 2020. p. 1–7.
- MITCHELL, T. *Machine Learning*. [S.l.]: McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673.
- MOHANDÉS, M.; DERICHE, M.; ALIYU, S. O. Classifiers combination techniques: A comprehensive review. *IEEE Access*, v. 6, p. 19626–19639, 2018.
- MORAES, R. M.; SOARES, E. A. d. M. G.; MACHADO, L. d. S. A double weighted fuzzy gamma naive Bayes classifier. *Journal of Intelligent & Fuzzy Systems*, v. 38, n. 1, p. 577–588, 2020.
- MOYANO, J. M. et al. Review of ensembles of multi-label classifiers: models, experimental study and prospects. *Information Fusion*, v. 44, p. 33–45, 2018.
- NABIPOUR, M. et al. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access*, v. 8, p. 150199–150212, 2020.
- NEMENYI, P. B. *Distribution-free multiple comparisons*. [S.l.]: Princeton University, 1963.
- NOLASCO, D. H. et al. Wavelet-fuzzy power quality diagnosis system with inference method based on overlap functions: Case study in an ac microgrid. *Engineering Applications of Artificial Intelligence*, v. 85, p. 284–294, 2019.
- PACHECO, A. G.; KROHLING, R. A. Aggregation of neural classifiers using Choquet integral with respect to a fuzzy measure. *Neurocomputing*, v. 292, p. 151–164, 2018.
- PAIVA, R. et al. Residuated implications derived from quasi-overlap functions on lattices. *International Journal of Approximate Reasoning*, v. 134, p. 95–110, 2021.
- PAIVA, R. et al. Lattice-valued overlap and quasi-overlap functions. *Information Sciences*, v. 562, p. 180–199, 2021.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

- PHAM, B. T. et al. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. *Journal of Geomatics*, v. 10, p. 71–79, 2016.
- PIZZI, N. J.; PEDRYCZ, W. Aggregating multiple classification results using fuzzy integration and stochastic feature selection. *International Journal of Approximate Reasoning*, v. 51, n. 8, p. 883–894, 2010.
- RIJSBERGEN, C. van. *Information Retrieval, 2nd ed.* [S.l.]: London: Butterworths, 1979.
- RISH, I. et al. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*. [S.l.: s.n.], 2001. v. 3, n. 22, p. 41–46.
- RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 2002.
- SÁ, L. R. de et al. New combining rules for spatial clustering methods using sigma-count for spatial epidemiology. In: *IEEE. 2020 International Conference on Decision Aid Sciences and Application (DASA)*. [S.l.], 2020. p. 794–800.
- SCHCLAR, A. et al. Ensemble methods for improving the performance of neighborhood-based collaborative filtering. In: *Proceedings of the third ACM Conference on Recommender systems*. [S.l.: s.n.], 2009. p. 261–264.
- SESMA-SARA, M. et al. Strengthened ordered directionally monotone functions. links between the different notions of monotonicity. *Fuzzy Sets and Systems*, v. 357, p. 151–172, 2019.
- SHIEH, J.-I.; WU, H.-H.; LIU, H.-C. Applying a complexity-based Choquet integral to evaluate students' performance. *Expert Systems with Applications*, v. 36, n. 3, p. 5100–5106, 2009.
- SILVA, I. A. da et al. Pseudo-uninorms and Atanassov's intuitionistic pseudo-uninorms. *Journal of Intelligent & Fuzzy Systems*, v. 29, n. 1, p. 267–281, 2015.
- SKLAR, M. Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, v. 8, p. 229–231, 1959.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, v. 45, n. 4, p. 427–437, 2009.
- SU, Y.; WANG, Z. Pseudo-uninorms and coimplications on a complete lattice. *Fuzzy Sets and Systems*, v. 224, p. 53–62, 2013.
- TAKÁČ, Z. et al. Discrete IV dG-Choquet integrals with respect to admissible orders. *Fuzzy Sets and Systems*, p. In press, 2021.
- TORRA, V.; NARUKAWA, Y. *Modeling decisions: information fusion and aggregation operators*. [S.l.]: Springer Science & Business Media, 2007.
- WOLOSZYNSKI, T.; KURZYNSKI, M. A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*, v. 44, n. 10-11, p. 2656–2668, 2011.

YANG, L. Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering*, v. 15, p. 4266–4270, 2011.

ZHOU, Z.-H. *Ensemble methods: foundations and algorithms*. [S.l.]: Chapman and Hall/CRC, 2019.

APÊNDICE A - Resultados das classificações para cada par de combinações com a AC-Integral utilizando a medida uniforme

Tabela 16: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^1 com os outros métodos

Base	$O_{PO}^1-O_{PO}^1$	$O_{PO}^1-O_{PO}^2$	$O_{PO}^1-O_{PO}^3$	$O_{PO}^1-O_{PO}^4$	$O_{PO}^1-O_Q^1$	$O_{PO}^1-O^1$	$O_{PO}^1-C^1$	$O_{PO}^1-O^2$	$O_{PO}^1-O_Q^2$
M1	91.87	92.57	92.57	93.73	93.04	91.64	93.03	92.80	92.57
MF	83.35	83.25	83.25	83.50	83.45	83.50	83.25	83.35	83.10
ST	84.45	84.40	84.45	84.30	84.45	84.50	84.30	84.35	84.55
SB	92.85	92.65	92.63	92.85	92.81	92.81	92.85	92.81	92.26
GC	74.70	74.90	75.00	74.50	74.50	75.30	74.70	74.90	73.70
OR	98.38	98.38	98.38	98.45	98.45	98.40	98.47	98.45	98.56
C9	94.63	95.00	95.09	94.35	94.17	94.72	94.35	94.26	93.15
SP	95.20	95.17	95.14	95.14	95.11	95.17	95.14	95.17	94.95
IR	95.33	96.00	96.00	95.33	95.33	96.00	95.33	95.33	95.33
HS	84.81	84.07	83.70	83.33	83.70	83.70	83.70	83.33	83.33
WR	55.22	56.41	56.54	55.85	55.54	55.91	55.54	55.47	53.72
TR	69.24	68.04	70.44	66.57	65.63	67.37	66.84	66.84	64.56
BU	74.20	73.33	73.62	72.46	72.75	73.91	72.17	72.17	71.01
HA	69.60	70.58	70.91	68.61	68.29	69.60	68.61	68.29	68.61
BS	77.28	79.84	80.16	78.08	77.60	79.04	78.56	78.56	72.96
BC	96.20	96.35	96.35	96.20	96.20	96.20	96.20	96.20	96.05
GL	93.42	93.44	93.44	92.49	92.49	93.44	92.97	92.97	93.44
IO	90.90	90.90	90.61	90.90	90.90	90.90	90.90	90.90	90.90
CE	79.81	80.85	80.85	80.62	80.74	80.45	80.68	80.68	78.71
NU	71.17	71.30	71.21	71.17	71.34	71.13	71.23	71.20	69.82
WA	85.12	85.64	85.66	84.84	84.72	85.50	84.98	85.04	84.16
VE	77.30	78.72	78.60	77.90	77.66	78.72	78.02	78.13	77.42
WW	51.02	52.12	52.49	50.92	51.00	51.51	51.10	51.21	49.49

Tabela 17: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^2 com os outros métodos

Base	$O_{PO}^2-O_{PO}^1$	$O_{PO}^2-O_{PO}^2$	$O_{PO}^2-O_{PO}^3$	$O_{PO}^2-O_{PO}^4$	$O_{PO}^2-O_Q^1$	$O_{PO}^2-O^1$	$O_{PO}^2-C^1$	$O_{PO}^2-O^2$	$O_{PO}^2-O_Q^2$
M1	90.94	90.94	91.64	90.48	93.96	91.40	93.03	92.80	83.99
MF	83.35	83.30	83.40	83.65	83.80	83.40	83.85	83.80	82.10
ST	84.35	84.30	84.40	84.25	84.20	84.30	84.05	84.05	83.00
SB	92.70	92.70	92.76	92.65	92.81	92.65	92.63	92.70	92.33
GC	74.90	75.00	75.00	75.30	74.70	75.00	75.40	75.20	74.80
OR	98.38	98.36	98.38	98.35	98.43	98.36	98.40	98.36	98.27
C9	95.00	94.91	94.81	95.19	94.17	94.91	94.63	94.91	92.31
SP	95.11	95.14	95.11	95.05	95.14	95.14	95.17	95.20	94.86
IR	96.00	96.00	96.00	96.00	95.33	96.00	96.00	96.00	96.00
HS	84.07	84.07	84.07	84.07	82.22	83.70	82.59	82.96	83.70
WR	56.35	56.35	55.60	55.91	53.41	56.47	53.91	54.60	50.66
TR	70.31	70.44	67.77	70.44	63.89	71.11	64.16	64.16	70.17
BU	73.33	73.62	73.62	73.91	69.57	73.91	69.28	69.86	73.33
HA	70.58	70.91	69.93	69.93	68.61	70.91	68.29	67.96	69.93
BS	79.68	79.68	79.52	79.84	77.12	80.16	77.76	78.24	74.72
BC	96.20	96.20	96.20	96.49	96.20	96.35	96.49	96.49	96.34
GL	93.44	93.44	92.98	92.97	92.02	93.44	91.57	92.50	92.03
IO	90.90	90.90	90.90	90.33	90.61	90.61	90.33	90.61	90.61
CE	80.68	80.50	80.45	80.79	80.74	80.85	80.85	81.03	78.25
NU	71.12	71.04	71.07	71.21	71.70	71.21	71.74	71.57	69.91
WA	85.60	85.62	85.50	85.76	84.08	85.60	84.64	84.84	85.12
VE	78.49	78.25	78.49	78.49	78.60	78.72	78.49	78.37	75.06
WW	51.98	52.04	51.80	51.39	50.63	52.67	51.00	51.29	43.84

Tabela 18: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^3 com os outros métodos

Base	$O_{PO}^3-O_{PO}^1$	$O_{PO}^3-O_{PO}^2$	$O_{PO}^3-O_{PO}^3$	$O_{PO}^3-O_{PO}^4$	$O_{PO}^3-O_Q^1$	$O_{PO}^3-O^1$	$O_{PO}^3-C^1$	$O_{PO}^3-O^2$	$O_{PO}^3-O_Q^2$
M1	91.17	92.10	90.94	92.34	93.96	91.40	93.03	93.03	90.02
MF	83.40	83.50	83.30	83.40	83.80	83.45	83.85	83.75	81.90
ST	84.35	84.30	84.25	84.25	84.25	84.20	84.15	84.05	83.25
SB	92.70	92.74	92.74	92.72	92.85	92.61	92.81	92.81	92.28
GC	75.00	75.40	75.10	75.10	74.40	75.20	75.60	75.70	74.70
OR	98.38	98.35	98.38	98.36	98.47	98.35	98.38	98.38	98.27
O_Q^2	95.00	95.00	94.91	95.28	94.17	95.00	94.72	94.81	92.13
SP	95.11	95.14	95.20	95.05	95.14	95.11	95.20	95.20	94.80
IR	96.00	96.00	96.00	96.00	95.33	96.00	95.33	95.33	96.00
HS	83.70	84.07	84.07	84.07	82.59	83.33	82.96	82.96	83.70
WR	56.41	56.22	56.16	55.72	53.41	56.47	53.85	53.85	48.16
TR	70.31	70.97	70.44	67.64	63.89	71.24	63.63	63.76	67.76
BU	73.33	73.91	73.33	73.91	69.86	73.62	69.28	69.57	73.04
HA	70.91	70.91	70.91	70.25	67.95	70.91	68.61	68.29	68.29
BS	79.68	80.16	79.52	79.20	76.80	80.32	77.76	77.92	72.96
BC	96.35	96.49	96.20	96.49	96.20	96.49	96.34	96.34	96.20
GL	93.44	93.91	94.37	93.44	92.02	93.44	90.62	92.02	92.48
IO	90.90	90.90	90.90	90.61	90.90	90.90	90.33	90.61	90.61
CE	80.56	80.56	80.62	80.91	80.79	80.56	80.85	80.97	77.78
NU	71.10	71.20	71.05	71.25	71.72	71.18	71.67	71.81	69.66
WA	85.64	85.64	85.60	85.62	84.12	85.74	84.30	84.48	84.64
VE	78.49	78.72	78.01	78.72	78.60	78.60	78.13	78.13	73.87
WW	52.14	50.96	51.61	51.90	50.39	52.78	50.47	50.78	41.26

Tabela 19: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^4 com os outros métodos

Base	$O_{PO}^4-O_{PO}^1$	$O_{PO}^4-O_{PO}^2$	$O_{PO}^4-O_{PO}^3$	$O_{PO}^4-O_{PO}^4$	$O_{PO}^4-O_Q^1$	$O_{PO}^4-O^1$	$O_{PO}^4-C^1$	$O_{PO}^4-O^2$	$O_{PO}^4-O_Q^2$
M1	93.26	92.33	92.57	92.57	93.73	94.43	93.26	93.26	94.65
MF	83.35	83.40	83.45	83.30	83.75	83.55	83.30	83.40	81.50
ST	84.30	84.00	84.30	84.50	84.30	84.30	84.45	84.45	82.65
SB	92.81	92.70	92.76	92.85	92.72	92.72	92.81	92.78	92.44
GC	74.40	74.70	75.10	74.90	75.30	73.90	75.30	75.30	75.10
OR	98.45	98.40	98.42	98.40	98.47	98.42	98.35	98.35	98.01
O_Q^2	94.35	94.54	94.63	94.26	94.26	94.07	94.81	94.91	92.04
SP	95.14	95.08	95.11	95.14	95.11	95.11	95.05	95.08	94.39
IR	95.33	96.00	95.33	95.33	95.33	95.33	96.00	96.00	96.00
HS	83.70	84.07	84.44	83.33	84.07	83.33	83.33	83.33	84.07
WR	55.91	55.60	55.54	55.29	55.66	54.72	56.22	56.22	54.66
TR	66.57	66.84	66.57	66.84	66.57	64.70	67.50	67.50	67.77
BU	72.46	74.20	72.75	72.17	72.75	73.04	72.46	73.04	73.33
HA	68.94	69.27	69.60	68.61	68.28	67.63	70.25	70.25	70.25
BS	78.08	77.76	77.44	78.24	79.04	76.96	79.84	79.84	73.92
BC	96.20	96.35	96.35	96.20	96.05	96.20	96.20	96.20	96.20
GL	92.49	92.03	91.55	92.49	92.97	92.02	93.44	93.44	93.42
IO	90.90	90.90	90.90	90.90	90.90	90.90	90.90	91.18	90.33
CE	80.62	80.39	80.50	80.62	80.97	80.68	80.85	80.97	77.43
NU	71.19	71.25	71.31	71.14	71.52	71.32	71.57	71.53	69.51
WA	84.88	85.10	84.84	85.04	84.92	84.42	85.62	85.62	85.56
VE	78.02	77.78	78.37	78.02	78.13	77.54	78.60	78.84	75.53
WW	51.00	51.53	50.82	51.16	51.35	50.90	51.78	51.80	47.82

Tabela 20: Resultados da aplicação da AC-Integral utilizando a função O_Q^1 com os outros métodos

Base	$O_Q^1-O_{PO}^1$	$O_Q^1-O_{PO}^2$	$O_Q^1-O_{PO}^3$	$O_Q^1-O_{PO}^4$	$O_Q^1-O_Q^1$	$O_Q^1-O^1$	$O_Q^1-C^1$	$O_Q^1-O^2$	$O_Q^1-O_Q^2$
M1	93.27	93.97	93.97	93.04	92.33	93.50	93.50	93.50	93.73
MF	83.40	83.45	83.45	83.35	83.50	83.50	83.55	83.45	83.65
ST	84.35	84.50	84.50	84.35	84.50	84.35	84.50	84.40	84.30
SB	92.72	92.83	92.85	92.76	92.83	92.74	92.70	92.78	92.76
GC	74.40	73.90	73.90	74.60	75.50	74.10	73.80	73.60	74.80
OR	98.45	98.45	98.45	98.43	98.38	98.43	98.43	98.43	98.45
O_Q^2	94.26	94.07	94.26	94.17	94.44	94.07	94.17	94.07	94.26
SP	95.14	95.11	95.14	95.08	95.08	95.14	95.11	95.11	95.14
IR	95.33	95.33	95.33	95.33	96.00	95.33	95.33	95.33	95.33
HS	83.70	83.33	83.33	83.70	83.70	83.70	83.33	82.96	84.07
WR	55.72	54.22	54.41	55.72	55.66	55.35	54.54	54.35	55.97
TR	66.30	62.96	62.82	66.17	67.10	65.10	64.83	64.70	66.84
BU	72.75	72.17	71.59	72.75	73.62	73.04	72.17	72.17	72.75
HA	68.61	65.99	65.99	68.61	69.60	68.29	67.63	66.65	69.26
BS	77.92	76.64	76.80	77.76	79.20	77.44	76.96	76.64	79.20
BC	96.20	96.20	96.20	96.20	96.05	96.20	96.20	96.20	96.20
GL	92.49	91.54	91.54	92.49	93.44	92.49	91.54	91.54	92.49
IO	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90	90.90
CE	80.68	80.45	80.56	80.68	80.45	80.68	80.51	80.51	80.79
NU	71.32	71.47	71.46	71.31	71.08	71.33	71.38	71.43	70.96
WA	84.78	84.26	84.32	84.74	85.26	84.56	84.46	84.44	84.92
VE	77.90	77.31	77.31	77.78	78.37	77.78	77.66	77.31	78.13
WW	50.96	50.37	50.47	51.04	51.47	50.96	50.65	50.67	51.23

Tabela 21: Resultados da aplicação da AC-Integral utilizando a função O^1 com os outros métodos

Base	$O^1-O_{PO}^1$	$O^1-O_{PO}^2$	$O^1-O_{PO}^3$	$O^1-O_{PO}^4$	$O^1-O_Q^1$	O^1-O^1	O^1-C^1	O^1-O^2	$O^1-O_Q^2$
M1	92.33	90.25	89.32	93.96	94.66	92.33	93.50	93.50	78.47
MF	83.35	82.75	83.00	83.80	83.60	83.45	83.80	83.75	81.45
ST	84.40	84.05	84.15	83.95	84.35	84.40	84.50	84.40	83.40
SB	92.83	92.57	92.54	92.87	92.87	92.85	92.70	92.81	91.41
GC	75.40	75.20	75.00	74.70	74.60	75.50	75.30	75.30	72.70
OR	98.38	98.36	98.36	98.40	98.49	98.38	98.45	98.40	97.81
O_Q^2	94.72	95.09	94.91	94.72	94.17	94.81	94.35	94.63	90.28
SP	95.11	95.05	94.95	95.20	95.14	95.14	95.05	95.08	93.95
IR	96.00	96.00	96.00	95.33	95.33	96.00	96.00	96.00	94.67
HS	83.70	83.33	83.70	82.96	83.70	84.07	83.70	83.70	78.89
WR	55.66	57.85	58.29	53.47	55.60	55.35	55.91	56.10	50.47
TR	67.10	71.64	72.04	63.09	65.64	66.97	66.84	66.84	58.81
BU	73.62	72.46	72.75	69.86	72.46	73.62	73.33	73.04	64.35
HA	69.60	71.57	71.57	68.29	67.96	69.60	68.61	68.28	66.98
BS	78.72	80.80	80.32	75.84	77.92	78.88	79.20	79.36	64.96
BC	96.05	96.64	96.64	96.20	96.20	96.05	96.20	96.20	95.91
GL	93.44	93.44	93.44	92.02	92.49	93.43	92.50	92.50	87.79
IO	90.90	90.90	90.33	91.18	90.61	90.90	90.61	90.90	91.47
CE	80.22	80.79	80.68	80.39	80.85	80.22	80.85	80.68	75.75
NU	71.06	72.13	71.97	72.24	71.44	71.10	71.52	71.49	68.24
WA	85.28	86.02	86.04	83.96	84.58	85.16	84.94	85.00	80.58
VE	78.01	79.31	79.08	78.13	77.90	78.01	78.37	78.25	72.46
WW	51.21	52.94	53.14	50.49	50.82	51.31	51.43	51.65	45.92

Tabela 22: Resultados da aplicação da AC-Integral utilizando a função C^1 com os outros métodos

Base	$C^1-O_{PO}^1$	$C^1-O_{PO}^2$	$C^1-O_{PO}^3$	$C^1-O_{PO}^4$	$C^1-O_Q^1$	C^1-O^1	C^1-C^1	C^1-O^2	$C^1-O_Q^2$
M1	92.57	94.43	94.43	92.57	91.40	93.03	91.64	91.64	93.50
MF	83.30	83.60	83.65	83.35	83.35	83.25	83.50	83.30	83.10
ST	84.40	84.30	84.35	84.40	84.10	84.35	84.20	84.20	84.30
SB	92.85	92.81	92.81	92.83	92.61	92.87	92.81	92.81	92.70
GC	74.90	73.70	73.90	75.40	75.00	74.60	75.30	75.20	75.20
OR	98.43	98.43	98.43	98.43	98.36	98.45	98.40	98.45	98.54
O_Q^2	94.26	94.17	94.17	94.35	95.00	94.26	94.81	94.63	94.44
SP	95.17	95.11	95.11	95.20	95.02	95.17	95.05	95.08	95.08
IR	95.33	96.00	95.33	95.33	96.00	95.33	95.33	95.33	96.00
HS	83.33	82.96	82.96	83.70	82.59	83.70	83.70	83.70	84.07
WR	55.72	54.53	54.66	55.72	56.35	55.29	55.66	55.85	55.85
TR	67.11	62.96	62.96	67.11	66.29	66.84	70.17	67.77	67.51
BU	72.46	72.17	72.17	73.33	70.72	72.17	73.33	73.62	73.04
HA	68.94	66.65	66.65	69.27	70.90	68.29	69.60	69.60	69.92
BS	78.72	76.80	76.80	78.88	80.96	78.72	79.36	79.36	79.68
BC	96.20	96.20	96.20	96.20	96.49	96.20	96.20	96.20	96.20
GL	92.97	91.54	91.54	92.97	92.50	92.49	92.97	92.97	92.96
IO	90.90	90.90	90.90	90.90	89.75	90.90	90.90	90.90	90.90
CE	80.62	80.68	80.68	80.56	80.91	80.68	80.39	80.39	80.04
NU	71.14	71.42	71.41	71.14	72.28	71.20	70.93	71.04	70.56
WA	85.06	84.48	84.42	85.12	85.72	84.96	85.36	85.36	85.52
VE	78.25	77.31	77.54	78.25	78.84	78.13	78.13	78.25	78.72
WW	51.08	50.51	50.61	51.21	52.61	50.94	52.06	51.72	51.45

Tabela 23: Resultados da aplicação da AC-Integral utilizando a função O^2 com os outros métodos

Base	$O^2-O_{PO}^1$	$O^2-O_{PO}^2$	$O^2-O_{PO}^3$	$O^2-O_{PO}^4$	$O^2-O_Q^1$	O^2-O^1	O^2-C^1	O^2-O^2	$O^2-O_Q^2$
M1	92.57	94.43	94.43	92.57	91.64	93.03	91.64	91.64	93.50
MF	83.30	83.60	83.65	83.45	83.45	83.35	83.60	83.40	83.10
ST	84.45	84.35	84.35	84.40	84.25	84.35	84.10	84.30	84.20
SB	92.85	92.81	92.81	92.83	92.63	92.85	92.83	92.76	92.68
GC	74.90	73.80	73.90	75.40	75.20	74.60	75.30	75.40	75.20
OR	98.43	98.47	98.45	98.42	98.38	98.45	98.36	98.38	98.51
O_Q^2	94.26	94.26	94.26	94.35	95.00	94.26	94.81	94.54	94.63
SP	95.17	95.11	95.11	95.17	95.05	95.17	95.02	95.05	95.08
IR	95.33	95.33	96.00	95.33	96.00	95.33	96.00	95.33	96.00
HS	83.33	82.96	82.96	83.70	82.59	83.70	84.44	84.07	84.07
WR	55.66	54.28	54.41	55.72	56.35	55.35	55.85	55.79	55.79
TR	67.11	64.43	64.16	67.11	65.49	66.84	70.17	67.77	67.37
BU	72.46	71.88	71.88	73.62	70.72	72.46	73.62	73.91	73.04
HA	68.94	66.65	66.65	69.27	70.25	68.29	69.93	69.60	69.92
BS	78.72	76.80	76.80	79.04	80.64	78.72	79.36	79.04	80.00
BC	96.20	96.20	96.20	96.20	96.49	96.20	96.20	96.20	96.20
GL	92.49	91.54	91.54	92.97	92.50	92.49	92.50	92.97	92.96
IO	90.90	90.90	90.90	90.90	89.75	90.90	90.90	90.90	90.90
CE	80.62	80.56	80.62	80.62	80.91	80.62	80.45	80.39	80.10
NU	71.13	71.37	71.37	71.16	72.38	71.18	71.06	70.98	70.56
WA	85.08	84.48	84.46	85.12	85.62	84.98	85.36	85.36	85.52
VE	78.25	77.66	77.54	78.25	79.08	78.13	78.01	78.37	78.60
WW	51.02	50.61	50.57	51.27	52.14	50.94	51.84	51.43	51.49

Tabela 24: Resultados da aplicação da AC-Integral utilizando a função O_Q^2 com os outros métodos

Base	$O_Q^2-O_{PO}^1$	$O_Q^2-O_{PO}^2$	$O_Q^2-O_{PO}^3$	$O_Q^2-O_{PO}^4$	$O_Q^2-O_Q^1$	$O_Q^2-O^1$	$O_Q^2-C^1$	$O_Q^2-O^2$	$O_Q^2-O_Q^2$
M1	92.10	80.78	81.94	79.85	91.87	92.57	89.09	88.86	92.33
MF	83.05	80.65	80.70	81.50	83.55	83.05	82.45	82.15	83.45
ST	84.55	82.30	81.60	83.00	84.60	84.50	84.25	84.05	84.40
SB	92.52	90.94	90.85	91.33	92.57	91.94	92.37	92.35	92.85
GC	74.50	71.50	70.40	72.00	75.30	73.40	75.40	75.40	75.50
OR	98.51	97.95	97.94	98.65	98.56	98.52	98.58	98.56	98.38
O_Q^2	93.89	87.78	88.24	90.56	94.07	91.76	92.78	92.78	94.81
SP	94.98	93.48	93.61	94.29	94.83	94.80	94.92	94.86	95.14
IR	95.33	94.67	94.67	94.00	95.33	95.33	96.00	96.00	96.00
HS	83.70	80.00	79.63	79.63	83.33	81.48	83.33	83.33	84.07
WR	54.22	44.97	44.15	48.28	55.85	52.66	56.66	56.41	55.35
TR	65.10	64.02	58.14	61.48	67.24	61.48	67.64	67.77	66.97
BU	71.88	68.99	64.64	68.70	71.88	71.30	73.04	73.04	73.62
HA	68.61	65.34	64.68	66.64	68.61	65.66	70.91	70.58	69.60
BS	74.40	62.40	61.60	65.44	78.24	70.24	76.48	76.32	78.88
BC	96.20	95.76	95.47	95.47	96.20	96.05	96.35	96.20	96.05
GL	93.44	92.46	91.98	93.90	92.50	93.43	92.98	92.98	93.43
IO	90.90	90.61	90.33	90.32	91.18	90.33	90.61	90.61	90.90
CE	79.69	75.64	75.64	75.29	80.45	77.84	78.13	78.25	80.22
NU	70.42	66.05	66.29	68.07	71.06	68.67	69.71	69.85	71.10
WA	84.64	80.48	79.20	81.28	84.92	83.42	85.34	85.42	85.16
VE	77.42	70.56	69.97	73.64	78.01	76.95	78.25	78.13	78.01
WW	49.82	38.02	38.69	42.61	51.45	48.69	51.76	51.12	51.31

APÊNDICE B - Resultados das classificações para cada par de combinações com a AC-Integral utilizando a medida ponderada

Tabela 25: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^1 com os outros métodos

Base	$O_{PO}^1-O_{PO}^1$	$O_{PO}^1-O_{PO}^2$	$O_{PO}^1-O_{PO}^3$	$O_{PO}^1-O_{PO}^4$	$O_{PO}^1-O_Q^1$	$O_{PO}^1-O^1$	$O_{PO}^1-C^1$	$O_{PO}^1-O^2$	$O_{PO}^1-O_Q^2$
MI	89.33	89.1	89.1	91.19	91.89	89.33	90.26	90.26	94.66
MF	83.6	83.15	83.35	83.55	83.7	83.35	83.65	83.5	83.2
ST	84.3	84.35	84.3	84.45	84.5	84.3	84.55	84.5	84.5
SB	92.81	92.72	92.76	92.78	92.87	92.72	92.83	92.81	92.35
GC	75.5	75.2	75.4	75.8	75.3	75.7	75.7	75.6	74.4
OR	98.52	98.65	98.63	98.61	98.61	98.56	98.59	98.61	98.56
C9	95.65	95.46	95.37	95.37	94.91	95.37	95.37	95.46	93.43
SP	95.39	95.17	95.14	95.45	95.42	95.39	95.42	95.42	95.08
IR	96	96	96	95.33	95.33	96	95.33	95.33	95.33
HS	84.07	82.96	83.33	82.96	83.7	82.96	83.7	83.33	83.33
WR	54.97	55.66	56.1	55.54	55.41	55.72	55.6	55.53	53.85
TR	73.51	72.71	72.71	69.38	67.64	70.31	68.31	68.31	65.23
BU	73.91	74.49	74.49	73.62	74.49	74.2	74.49	74.78	73.04
HA	74.84	74.51	74.51	73.86	71.23	74.84	73.53	73.53	68.94
BS	94.08	87.36	86.88	89.76	85.6	88.16	86.72	86.72	75.2
BC	96.2	96.35	96.49	96.34	96.2	96.2	96.34	96.34	96.05
GL	89.65	92.02	92.02	90.12	91.06	92.47	91.54	91.54	92.96
IO	91.75	90.04	90.04	91.47	90.9	90.9	90.9	90.9	90.9
CE	79.29	80.1	79.98	80.22	80.04	79.64	80.1	79.98	78.71
NU	71.28	70.52	70.32	70.57	70.46	70.13	70.5	70.49	69.02
WA	86.76	86.24	86.28	85.84	85.36	86.24	85.78	85.8	84.54
VE	81.56	80.02	80.14	79.32	78.72	79.55	79.55	79.79	77.54
WW	46.76	50.96	51.18	49.72	50.21	50.29			

Tabela 26: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^2 com os outros métodos

Base	$O_{PO}^2-O_{PO}^1$	$O_{PO}^2-O_{PO}^2$	$O_{PO}^2-O_{PO}^3$	$O_{PO}^2-O_{PO}^4$	$O_{PO}^2-O_Q^1$	$O_{PO}^2-O^1$	$O_{PO}^2-C^1$	$O_{PO}^2-O^2$	$O_{PO}^2-O_Q^2$
M1	89.1	89.8	89.8	89.1	96.98	89.1	95.59	95.59	89.57
MF	83.45	83.9	83.75	83.6	83.6	83.3	83.65	83.5	82.05
ST	84.3	83.85	84.15	84	84.15	84.1	84.15	83.9	83.4
SB	92.85	92.87	92.85	92.89	92.83	92.78	92.78	92.63	92.63
GC	75.6	75.3	75.5	76.1	74.8	75.1	75	75	75.7
OR	98.67	98.63	98.65	98.68	98.58	98.74	98.43	98.42	98.45
C9	95.56	95.65	95.56	95.46	94.91	95.65	95.37	95.37	94.63
SP	95.3	95.14	95.24	95.36	95.36	95.36	95.11	95.2	95.14
IR	96	96	96	96	95.33	96	96	96	96
HS	83.33	84.44	84.44	83.7	82.59	83.7	82.22	82.22	83.7
WR	56.04	55.85	56.35	55.66	53.22	55.91	53.97	54.28	50.6
TR	72.98	73.65	73.38	73.38	63.76	73.38	61.89	64.29	72.98
BU	73.04	74.49	73.62	73.91	70.14	73.91	68.99	69.28	74.49
HA	74.84	74.84	75.49	75.82	68.29	74.18	68.61	69.27	72.55
BS	92	93.28	93.28	93.12	79.04	92.8	78.72	79.68	80.96
BC	96.35	96.49	96.2	96.64	96.2	96.49	96.49	96.49	96.64
GL	91.99	91.04	90.58	91.52	92.02	91.52	93.44	93.44	91.53
IO	90.61	91.18	91.47	90.33	90.61	90.33	91.18	91.18	90.9
CE	80.22	80.16	80.33	80.22	80.45	80.22	80.1	79.81	78.88
NU	70.66	72	72.13	71.47	71.17	71.67	72.36	72.27	70.08
WA	86.6	86.8	86.8	86.76	83.84	86.74	84.54	84.98	86.44
VE	81.21	81.8	81.92	81.56	77.9	81.8	77.9	78.25	77.43
WW	50.45	47.27	47.43	49.82	50.49	50.1	50.82	50.65	44.88

Tabela 27: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^3 com os outros métodos

Base	$O_{PO}^3-O_{PO}^1$	$O_{PO}^3-O_{PO}^2$	$O_{PO}^3-O_{PO}^3$	$O_{PO}^3-O_{PO}^4$	$O_{PO}^3-O_Q^1$	$O_{PO}^3-O^1$	$O_{PO}^3-C^1$	$O_{PO}^3-O^2$	$O_{PO}^3-O_Q^2$
M1	89.1	89.57	89.8	89.33	96.98	88.87	95.59	95.36	89.8
MF	83.45	83.7	83.8	83.45	83.75	83.1	83.6	83.6	82.25
ST	84.15	83.95	83.85	84.3	84.15	84.1	84.1	84.15	83.25
SB	92.81	92.94	92.89	92.89	92.96	92.81	92.78	92.81	92.5
GC	75.6	75.7	75.5	75.8	74.5	75.2	74.9	75.2	75.8
OR	98.68	98.63	98.63	98.68	98.59	98.72	98.47	98.43	98.51
C9	95.56	95.56	95.56	95.56	94.72	95.56	95.37	95.37	94.26
SP	95.3	95.11	95.17	95.24	95.3	95.33	95.17	95.11	95.17
IR	96	96	96	96	95.33	96	95.33	95.33	96
HS	83.7	84.44	84.44	84.07	82.96	83.7	82.22	82.22	83.7
WR	55.85	55.22	55.91	55.66	53.22	56.1	53.35	53.91	47.34
TR	73.11	74.05	73.51	71.77	63.49	73.38	61.63	64.16	69.64
BU	73.33	74.49	73.91	74.49	70.43	73.62	68.41	69.28	73.91
HA	74.84	75.49	74.84	75.49	68.61	74.18	67.96	67.96	70.91
BS	92	92.96	93.28	93.44	79.68	92.64	79.52	80.16	79.2
BC	96.49	96.49	96.49	96.64	96.2	96.49	96.34	96.34	96.2
GL	91.99	91.52	91.04	91.52	92.02	91.52	92.97	92.97	91.52
IO	90.61	91.18	91.18	90.33	90.9	90.33	91.18	91.18	90.9
CE	80.22	79.75	80.1	80.04	80.33	80.22	80.04	79.75	78.59
NU	70.62	72.22	71.99	71.13	71.17	71.65	72.38	72.25	69.98
WA	86.58	86.7	86.8	86.64	83.86	86.72	84.24	84.48	86.2
VE	81.45	81.33	81.68	80.73	77.07	82.04	77.42	77.54	76.6
WW	50.18	47.27	47.14	49.27	50.47	50.02	50.51	50.51	44.26

Tabela 28: Resultados da aplicação da AC-Integral utilizando a função O_{PO}^4 com os outros métodos

Base	$O_{PO}^4-O_{PO}^1$	$O_{PO}^4-O_{PO}^2$	$O_{PO}^4-O_{PO}^3$	$O_{PO}^4-O_{PO}^4$	$O_{PO}^4-O_Q^1$	$O_{PO}^4-O^1$	$O_{PO}^4-C^1$	$O_{PO}^4-O^2$	$O_{PO}^4-O_Q^2$
M1	89.57	89.8	90.03	89.33	94.43	90.73	93.96	93.73	94.2
MF	83.6	83.55	83.6	83.75	83.55	83.55	83.35	83.25	82.05
ST	84.4	84.4	84.45	84.4	84.35	84.4	84.35	84.4	83.2
SB	92.91	92.87	92.83	92.94	92.78	92.91	92.7	92.7	92.46
GC	75.8	74.9	75.1	75.7	75.4	75.8	75.4	75.6	75.4
OR	98.65	98.63	98.65	98.65	98.54	98.63	98.65	98.63	98.19
C9	95.46	95.46	95.46	95.83	95.19	95.37	95.37	95.46	93.52
SP	95.42	95.45	95.27	95.33	95.36	95.39	95.14	95.17	94.7
IR	95.33	96	95.33	95.33	96	95.33	96	96	96
HS	82.96	82.96	82.96	83.33	83.33	83.33	83.33	83.33	84.07
WR	55.6	55.41	55.47	55.6	55.04	54.91	56.22	56.29	52.97
TR	70.71	72.44	69.64	70.45	67.11	70.58	67.77	68.04	68.71
BU	73.62	73.91	74.2	72.75	71.59	74.2	72.46	72.46	74.49
HA	75.49	73.2	72.55	75.82	70.25	75.16	71.56	71.56	71.89
BS	92.8	89.76	90.88	93.92	82.24	93.44	82.56	83.36	78.72
BC	96.34	96.49	96.49	96.34	96.2	96.2	96.2	96.2	96.34
GL	91.04	90.59	90.58	89.65	92.02	89.18	92.5	92.5	92.97
IO	91.18	90.9	90.9	91.47	90.9	91.47	90.9	90.9	90.9
CE	80.22	79.93	79.64	79.98	80.62	80.22	80.56	80.51	76.86
NU	70.79	71.52	71.35	72.2	70.84	72.37	70.93	70.76	69.96
WA	86.34	86.04	85.92	86.72	85.08	86.24	85.64	85.78	85.8
VE	80.97	80.02	79.79	80.97	79.08	80.26	78.25	78.13	77.54
WW	48.63	49.98	48.96	47.14	51.27	47.59	52.21	51.92	44.59

Tabela 29: Resultados da aplicação da AC-Integral utilizando a função O_Q^1 com os outros métodos

Base	$O_Q^1-O_{PO}^1$	$O_Q^1-O_{PO}^2$	$O_Q^1-O_{PO}^3$	$O_Q^1-O_{PO}^4$	$O_Q^1-O_Q^1$	$O_Q^1-O^1$	$O_Q^1-C^1$	$O_Q^1-O^2$	$O_Q^1-O_Q^2$
M1	90.5	91.66	91.66	90.26	89.33	90.73	90.73	90.73	89.33
MF	83.65	83.55	83.55	83.7	83.75	83.65	83.8	83.7	83.4
ST	84.6	84.6	84.6	84.55	84.1	84.5	84.45	84.45	84.4
SB	92.83	92.85	92.85	92.83	92.85	92.81	92.94	92.87	92.87
GC	75.6	75	75.1	75.4	75.7	75.5	75.3	75.3	75.6
OR	98.68	98.7	98.68	98.68	98.65	98.68	98.7	98.72	98.74
C9	95.46	95.28	95.28	95.65	95.65	95.56	95.28	95.28	95.46
SP	95.42	95.39	95.33	95.45	95.2	95.49	95.36	95.33	95.3
IR	95.33	95.33	95.33	95.33	96	95.33	95.33	95.33	95.33
HS	83.7	83.7	83.7	83.7	83.7	83.7	83.7	83.7	82.96
WR	55.29	54.22	54.1	54.97	56.04	55.16	54.66	54.41	55.29
TR	68.58	66.3	66.3	68.04	73.38	68.31	69.38	68.98	69.64
BU	74.78	73.91	73.91	74.49	73.62	74.78	74.49	74.49	74.49
HA	73.2	71.23	71.56	73.2	74.84	73.2	74.51	74.83	75.17
BS	89.6	89.6	89.6	89.28	92.48	89.6	91.52	91.04	90.72
BC	96.2	96.35	96.35	96.2	96.2	96.2	96.2	96.2	96.2
GL	90.59	88.72	88.72	91.06	89.65	90.59	88.72	88.72	90.11
IO	91.47	91.47	91.47	91.18	91.47	91.47	91.47	91.47	91.47
CE	80.04	80.16	80.22	79.93	80.27	80.04	80.16	80.16	80.1
NU	70.38	70.56	70.56	70.35	72.38	70.51	70.51	70.45	69.63
WA	85.88	84.92	85	85.74	86.46	85.64	85.94	85.88	86.12
VE	79.08	78.96	78.96	79.32	81.68	79.08	79.91	79.79	79.79
WW	49.94	48.55	48.71	49.94	46.96	49.65	48.35	48.37	49.9

Tabela 30: Resultados da aplicação da AC-Integral utilizando a função O^1 com os outros métodos

Base	$O^1-O_{PO}^1$	$O^1-O_{PO}^2$	$O^1-O_{PO}^3$	$O^1-O_{PO}^4$	$O^1-O_Q^1$	O^1-O^1	O^1-C^1	O^1-O^2	$O^1-O_Q^2$
MI	89.33	88.87	88.87	91.19	95.12	89.33	94.66	94.66	89.79
MF	83.6	83.3	83.1	83.75	83.75	83.55	83.55	83.6	81.95
ST	84.35	84.2	83.85	84.35	84.55	84.25	84.5	84.45	83.35
SB	92.87	92.7	92.61	92.98	92.89	92.89	92.76	92.76	91.39
GC	75.5	75.2	75.6	76.1	74.9	75.6	75.7	75.6	72.2
OR	98.63	98.63	98.61	98.59	98.56	98.61	98.58	98.58	98.02
C9	95.46	95.46	95.56	95.37	94.81	95.46	95.19	95.28	91.67
SP	95.42	95.36	95.3	95.08	95.39	95.36	95.2	95.24	94.39
IR	96	96	96	95.33	95.33	96	96	96	95.33
HS	82.96	84.07	84.44	82.59	84.07	83.7	83.33	83.33	79.26
WR	55.66	57.6	57.47	53.1	54.85	55.6	55.16	55.41	48.41
TR	70.71	72.44	72.85	66.03	65.9	73.11	67.1	67.37	57.48
BU	72.17	73.33	73.33	71.01	72.75	73.33	72.46	72.75	67.25
HA	75.49	74.84	74.51	69.59	69.59	75.49	70.9	71.23	66.32
BS	93.44	85.44	86.4	91.04	83.36	94.08	82.88	83.2	68.96
BC	96.2	96.64	96.64	96.35	96.2	96.2	96.2	96.2	95.91
GL	91.52	93.43	92.96	89.66	92.02	90.11	92.97	92.49	89.21
IO	91.18	90.9	90.9	91.18	90.9	91.47	90.9	90.9	90.61
CE	80.1	79.81	79.64	79.06	80.45	79.93	80.56	80.62	75.52
NU	70.6	72.11	72.15	72.07	70.76	71.71	70.83	70.82	68.8
WA	86.54	86.34	86.38	85.56	84.98	86.74	85.54	85.58	82.12
VE	80.62	80.62	80.38	78.61	78.96	81.33	78.72	78.84	73.28
WW	48.82	52.35	52.33	47.53	50.67	47.1	51.16	50.94	43.18

Tabela 31: Resultados da aplicação da AC-Integral utilizando a função C^1 com os outros métodos

Base	$C^1-O_{PO}^1$	$C^1-O_{PO}^2$	$C^1-O_{PO}^3$	$C^1-O_{PO}^4$	$C^1-O_Q^1$	C^1-O^1	C^1-C^1	C^1-O^2	$C^1-O_Q^2$
M1	89.57	90.96	90.96	89.33	94.66	89.57	89.1	89.1	89.1
MF	83.45	83.45	83.35	83.45	83.45	83.4	84	83.6	83.35
ST	84.5	84.4	84.45	84.55	84.2	84.5	83.95	83.95	84.3
SB	92.89	92.89	92.85	92.87	92.57	92.89	92.91	92.94	92.68
GC	75.7	75.3	75.4	75.5	75	75.5	75.6	75.5	75.7
OR	98.67	98.68	98.68	98.67	98.52	98.67	98.7	98.7	98.74
C9	95.56	95.65	95.65	95.56	95.37	95.56	95.46	95.46	95.46
SP	95.39	95.39	95.36	95.49	95.17	95.42	95.02	95.2	95.14
IR	96	95.33	95.33	96	96	95.33	96	95.33	96
HS	83.33	84.07	84.07	82.59	82.59	82.96	84.07	84.44	83.33
WR	55.41	54.53	54.72	55.53	55.85	55.35	55.91	55.72	55.79
TR	69.78	67.64	67.64	72.71	67.1	69.78	73.91	74.05	70.71
BU	74.49	73.33	73.91	73.91	71.01	74.2	74.2	73.33	72.46
HA	74.51	72.88	73.2	74.51	70.25	74.18	74.51	75.16	75.82
BS	90.24	91.04	91.04	90.08	79.36	90.08	92.8	93.44	91.2
BC	96.34	96.35	96.35	96.34	96.49	96.34	96.34	96.34	96.34
GL	90.6	88.72	88.72	90.6	92.5	90.11	89.65	89.18	90.11
IO	90.9	91.47	91.47	90.9	90.04	91.47	91.47	91.47	90.9
CE	80.04	80.22	80.27	79.98	80.68	80.16	80.22	80.33	80.33
NU	70.53	70.94	71.02	70.53	72.11	70.74	72.34	72.39	70.02
WA	86.2	85.62	85.6	86.22	85.5	86.18	86.68	86.7	86.54
VE	79.91	79.79	79.79	79.79	78.61	79.91	81.56	81.56	80.97
WW	49.8	48.37	48.39	50.04	51.82	49.55	47.22	47.37	49.51

Tabela 32: Resultados da aplicação da AC-Integral utilizando a função O^2 com os outros métodos

Base	$O^2-O_{PO}^1$	$O^2-O_{PO}^2$	$O^2-O_{PO}^3$	$O^2-O_{PO}^4$	$O^2-O_Q^1$	O^2-O^1	O^2-C^1	O^2-O^2	$O^2-O_Q^2$
MI	89.57	90.96	91.19	89.33	94.66	89.57	89.1	89.1	89.33
MF	83.35	83.4	83.5	83.45	83.3	83.45	83.85	83.9	83.25
ST	84.35	84.35	84.4	84.55	84.2	84.5	84.25	84.1	84.3
SB	92.91	92.87	92.85	92.85	92.57	92.89	92.91	92.85	92.68
GC	75.7	75.4	75.5	75.5	75.3	75.5	75.1	75.3	75.7
OR	98.63	98.65	98.67	98.65	98.54	98.65	98.65	98.67	98.74
C9	95.56	95.56	95.65	95.56	95.19	95.46	95.56	95.56	95.56
SP	95.39	95.39	95.42	95.49	95.2	95.45	95.11	95.11	95.11
IR	96	95.33	95.33	96	96	95.33	96	96	96
HS	83.33	84.07	84.07	82.59	82.96	82.96	84.07	84.07	83.33
WR	55.47	54.66	54.6	55.47	55.66	55.28	56.03	55.97	55.85
TR	69.78	67.9	67.77	72.71	63.09	69.64	73.38	73.51	70.71
BU	74.49	73.91	74.49	73.91	70.72	74.2	74.2	74.49	72.46
HA	74.51	73.2	73.2	74.51	70.58	74.51	74.18	75.17	75.82
BS	90.24	91.2	91.2	90.24	78.24	90.24	92.8	92.96	91.68
BC	96.34	96.35	96.35	96.34	96.49	96.34	96.49	96.34	96.34
GL	90.6	88.72	88.72	90.6	92.5	90.11	90.58	90.11	90.11
IO	90.9	91.47	91.47	90.9	90.04	91.47	91.47	91.47	90.9
CE	80.04	80.04	80.1	79.87	80.56	80.16	80.16	80.22	80.33
NU	70.54	70.92	70.96	70.51	72.17	70.67	72.21	72.28	69.89
WA	86.18	85.72	85.72	86.22	85.32	86.16	86.58	86.74	86.52
VE	79.67	79.91	79.91	80.02	77.78	80.02	81.56	81.92	81.09
WW	49.78	48.35	48.37	50	52.02	49.55	47.78	47.27	49.37

Tabela 33: Resultados da aplicação da AC-Integral utilizando a função O_Q^2 com os outros métodos

Base	$O_Q^2-O_{PO}^1$	$O_Q^2-O_{PO}^2$	$O_Q^2-O_{PO}^3$	$O_Q^2-O_{PO}^4$	$O_Q^2-O_Q^1$	$O_Q^2-O^1$	$O_Q^2-C^1$	$O_Q^2-O^2$	$O_Q^2-O_Q^2$
M1	90.26	95.37	95.6	81.01	93.96	93.52	92.57	92.33	89.33
MF	83.35	80.95	81.5	81.65	83.6	83	82.85	82.4	83.55
ST	84.3	82.55	82	83.35	84.4	84.35	84.4	84.15	84.25
SB	92.52	91.13	91.18	91.7	92.65	92.15	92.26	92.22	92.89
GC	75.3	73.4	72.1	72.7	76.1	73.4	75.8	75.7	75.6
OR	98.65	98.49	98.45	98.75	98.75	98.59	98.74	98.7	98.61
C9	94.72	91.67	92.04	93.33	95.46	94.26	94.54	94.26	95.46
SP	95.27	93.79	93.92	94.36	95.11	95.02	94.98	94.92	95.36
IR	95.33	94.67	94.67	94	96	95.33	96	96	96
HS	84.07	81.11	81.11	81.48	83.7	82.96	82.96	83.33	83.7
WR	54.53	43.59	43.03	45.59	55.22	53.35	56.1	55.72	55.6
TR	68.04	65.35	60.54	64.03	67.91	64.56	68.97	69.24	73.11
BU	75.07	70.72	69.28	68.99	73.33	73.62	72.75	73.33	73.33
HA	71.89	67.3	65.66	68.61	72.22	69.6	74.18	73.86	75.49
BS	84.8	68	68.8	71.2	85.28	78.24	82.08	82.24	94.08
BC	96.2	95.76	95.91	95.62	96.2	96.05	96.35	96.35	96.2
GL	91.06	91.04	90.58	92.92	92.02	89.15	92.97	92.97	90.11
IO	91.47	90.9	90.61	90.9	91.18	91.18	90.61	90.61	91.47
CE	79.46	75.81	76.56	76.34	80.16	77.78	78.3	77.96	79.93
NU	69.73	65.69	66.11	66.77	69.73	68.89	68.02	67.99	71.71
WA	85.56	83.08	82.52	83.24	85.54	84.58	86.18	86.24	86.74
VE	78.96	74.58	75.18	72.69	78.96	77.66	79.55	79.67	81.33
WW	48.55	37.36	37.85	40.73	50.82	46.96	50.63	49.96	47.1

APÊNDICE C – Descrição das Bases de Dados

Em seguida serão descritas as bases de dados utilizadas neste trabalho.

- **Monk's Problems - M1**

A base de dados *Monk's Problems* é uma base de dados utilizada em uma das primeiras comparações entre algoritmos de aprendizagem. Utilizamos uma das três bases disponíveis com este conjunto de teste

- **Multiple Features - Fourier - MF**

Esta base consiste de uma base de dados referente a dígitos (0–9) que foram transformados em uma imagem binária. Neste trabalho, utilizaremos a base que contém como atributos 76 coeficientes de Fourier das formas dos dígitos

- **Spam - SB**

Esta base consiste de um conjunto de características extraídas de e-mails para verificar se determinado e-mail é um *spam* ou não. Possui duas classes, que indicam se o e-mail foi classificado ou não como spam.

- **CNAE-9 - C9**

Esta base contém descrições de empresas brasileiras que são divididas em 9 categorias. O número de atributos dessa classe é 857.

- **German Credit - GC**

Esta base contém informações de pessoas com relação a risco de crédito. Contém 20 atributos e duas classes, que indicam se o cliente é bom ou ruim.

- **Molecular Biology (Splice) - SP**

Esse conjunto de dados contém dados com informações relativas a moléculas de DNA. Possui 61 atributos e 3 classes.

●**Optical Recognition** - OP

Esta base contém uma imagem bitmap extraídas a partir de uma dígitos escritos a mão para visão computacional. Possui nove classes referentes aos dígitos (0–9)

●**Heart Statlog** - HS

Heart statlog é uma base originalmente com 270 instâncias e 13 atributos referentes a dados de doenças do coração em pacientes. Possui duas classes, que identificam se o paciente possui doença no coração ou não.

●**Wine Quality** - WR e WW

As bases *Wine Quality* se refere a duas bases diferentes com dados de dois tipos de vinho: tinto e branco com 6 e 7 classes, respectivamente, que indicam a qualidade do vinho em uma escala de 0 a 10 (nem todos os valores são utilizados). Essa base possui originalmente 4898 instâncias e 12 atributos para identificar o vinho.

●**Transfusion** - TR

A base de dados *Blood Transfusion* é uma base de dados de classificação com dados de doadores de sangue coletado em Taiwan. Essa base possui 748 instâncias e 5 atributos. Possui duas classes que indicam se a pessoa doou sangue em Março de 2007 ou não.

●**Bupa** - BU

A base *Liver Disorders*, também conhecida como Bupa, possui dados de 345 exames de sangue feitos em pacientes. Possui 2 classes (que identificam se o paciente bebe ou não) e 7 atributos.

●**Haberman** - HA

A base de dados *Haberman's Survival* contém dados de um estudo que foi conduzido sobre a taxa de sobrevivência de pacientes que realizaram cirurgia de câncer de mama. Possui dados de 306 pacientes, com 3 atributos e duas classes, que identificam se o paciente viveu em menos de 5 anos ou sobreviveu mais de 5 anos.

●**Balance Scale** - BS

A base de dados *Balance Scale* é uma base artificial que simula uma balança, com duas variáveis do lado esquerdo, duas do lado direito (totalizando 4 atributos) e 3 classes, que indicam se a balança irá pender para o lado esquerdo, direito, ou ficará balanceada. Possui 625 instâncias.

- **Iris - IR**

A base de dados Iris possui dados de 3 classes de flores, cada uma com 50 instâncias (totalizando 150), possuindo 4 atributos referente a flor. É uma das bases de dados mais encontrada na literatura em problemas de classificação.

- **Breast Cancer - BC**

Essa base possui dados características de células humanas encontradas em nódulos encontrados nos seios de pacientes. Possui 569 instâncias e 32 atributos, com duas classes que identificam se o nódulo é maligno ou benigno.

- **Glass - GL**

A base de dados *Glass* possui dados de 6 tipos de vidro e 214 instâncias, com 10 atributos que identificam as características do vidro da instância.

- **Ionosphere - IO**

Essa base contém dados eletromagnéticos coletados de 16 antenas. O objetivo dessa base é identificar se o sinal é bom ou ruim, totalizando duas classes. Essa base possui 351 instâncias e 34 atributos.

- **Car Evaluation - CE**

A base *Car Evaluation* possui dados de características de carros. Possui 4 classes que identificam se o valor sendo cobrado pelo carro é inaceitável, aceitável, bom ou muito bom. Possui dados de 1728 carros e 6 atributos.

- **Nursery - NU**

A base de dados *Nursery* contém dados de inscrições em um colégio de Enfermagem. Possui dados de 12960 inscrições e 8 atributos que identificam o aplicante. A base de dados não possui uma classe bem estabelecida, portanto foi escolhido o 8º atributo como classe.

- **Waveform - WA**

Essa base de dados possui dados gerados a partir de ondas, possui 3 classes referentes a 3 tipos de ondas, 5000 instâncias e 21 atributos.

- **Vehicle - VE**

A base de dados *Vehicle* possui dados de extração de silhuetas de objetos, possui 18 atributos que identificam o elemento. Possui 946 instâncias e 4 classes.