

RONAN ASSUMPÇÃO SILVA

**USO DA CENTRALIDADE DE REDES
COMPLEXAS NA COMBINAÇÃO E NA
SELEÇÃO DE CLASSIFICADORES**

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Doutor em Informática.

Curitiba

2018

RONAN ASSUMPÇÃO SILVA

**USO DA CENTRALIDADE DE REDES
COMPLEXAS NA COMBINAÇÃO E NA
SELEÇÃO DE CLASSIFICADORES**

Tese apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Doutor em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Alceu de Souza Britto Jr.
Co-orientadores: Prof. Dr. Fabricio Enembreck
e Prof. Dr. Robert Sabourin

Curitiba

2018

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Giovanna Carolina Massaneiro dos Santos – CRB 9/1911

S586u
2018

Silva, Ronan Assumpção
Uso da centralidade de redes complexas na combinação e na seleção de classificadores / Ronan Assumpção Silva ; orientador: Alceu de Souza Britto Junior; coorientadores: Fabricio Enembreck, Robert Sabourin. - 2018
105 f. : il. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2018
Bibliografia: f. 89-98

1. Processamento de dados. 2. Redes de computação. 3. Algoritmos.
4. Reconhecimento de padrões. I. Britto Junior, Alceu de Souza.
I. Enembreck, Fabricio. I. Sabourin, Robert. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática. III. Título.

CDD 22. ed. – 004

*Este trabalho é dedicado a minha esposa,
companheira fiel e motivadora,
que tem me inspirado a superar grandes desafios.*

Agradecimentos

Agradeço primeiramente a Deus, meu refúgio e minha fortaleza. Agradeço ao meu orientador, Dr. Alceu de Souza Britto Jr., que acreditou e investiu no meu potencial. Agradeço também ao Dr. Fabrício Enembreck, meu co-orientador, cuja sabedoria e conselhos enriqueceram por demais este trabalho. Meu muito obrigado também ao Dr. Robert Sabourin, meu co-orientador, por toda contribuição e por ter me recebido no período em que fiz pesquisa no Canadá. Aos demais membros da banca, Dr. Alessandro L. Koerich, Dr.^a Deborah R. Carvalho e Dr. Júlio Cesar Nievola, pelas contribuições neste trabalho, meu muito obrigado. Também agradeço aos professores: Dr.^a Alaine M. Guimarães, Dr.^a Andreia Malucelli, Dr. Edson Justino, Dr. Emerson Paraiso, Dr. Flávio Bortolozzi, Dr. Gleifer Vaz Alves, Dr. José Carlos F. da Rocha e Dr. Luiz S. de Oliveira.

Agradeço aos amigos do PPGIA: Alexandre Belarmino, Anderson Bertling, André Brun e Greicy Kiel, Andreia Marini, Bruno Souza, Cheila Cristina, Cleverton Vicentini, Denise Sato, Diogo Olsen, Edenilson José da Silva, Eduardo Viegas, Elias Carvalho, Erich Malinowski, Estefânia Fuzyi, Fabiano Utiyama, Flávia Beuting, Flávio de Almeida e Silva, Francis Baranoski, Gustavo Bonacina, Heitor Gomes, Irapuru Florido e Débora, Jackson Mallmann, Jean-Paul Barddal, Jhonatan Geremias, Jurandir dos Santos, Kelly Wiggers, Luiz Giovanini, Marcelo Pereira, Marcia Pascutti, Mariza Dosciatti, Marcos Monteiro Junior, Nicolas de Paula, Patrícia Antonioli, Rodolfo Botto, Rodrigo Siega, Rosana Lachowski, Sandoval Ruppel, Sediane Hernandez, Sidnei Schuindt, Vilmar Abreu, Viviane Dal Molin, Voncarlos Araújo e Wendel Goes.

Agradeço aos amigos do Canadá, da ÉTS e do LIVIA Lab: Alexandre R. de Mello e Jessica Freitag, Amin Abbasnejad, Ana C. Riekstin e Carlos E. Ladeira, Hiba Zakane, Hongliu Cao, Jonathan de Matos e família, Luiz G. Hafemann, Mirmohammad Saadati, Moslem Ouled Sghaier, Rafael Menelau, Rafooneh Jafarian Bahri, Sajjad Abdoli, Saypraseuth Mounsaveng e Thanh Le.

Não seria possível a realização deste trabalho sem suporte financeiro. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Código de Financiamento 88881.131663/2016-01. Agradeço a CAPES e também ao Instituto Federal do Paraná (IFPR), onde trabalho desde 2011.

Muito obrigado a minha família, em especial, minha esposa.

*“Ensina-nos a contar os nossos dias
para que o nosso coração alcance sabedoria.”*

Salmos 90:12

Resumo

O conjunto de classificadores é usado como alternativa aos sistemas monolíticos para tentar aumentar a precisão de um sistema de reconhecimento de padrões. Um Sistema de Múltiplos Classificadores típico pode ser dividido em três fases: Geração, Seleção e Fusão. Nesta última, geralmente, um critério para ponderar a decisão dos membros do conjunto gerado é adotado. Para tal, pode empregar alguma heurística como precisão ou confiança. Então, alguma regra ou fórmula deve usar o critério estabelecido para fornecer a influência do classificador e promover a competição dos classificadores no processo de rotulação. Este trabalho apresenta um novo método de fusão, a Fusão Baseada em Centralidade (*Centrality Based Fusion - CBF*), que utiliza o conceito de centralidade no contexto da Análise de Redes Sociais como um critério para a decisão do conjunto. Medidas de centralidade são aplicadas no contexto da SNA para medir a importância de cada pessoa dentro de uma rede social levando em conta as relações que cada pessoa tem com todas as outras. Então, a ideia é derivar a pontuação dos classificadores considerando a proeminência geral do classificador dentro da rede de classificadores, a qual reflete as relações pareadas dos classificadores. Nós hipotetizamos que pela análise da posição de um classificador na rede, isto poderia ser usado como a influência do classificador definindo o peso do voto dele para a fase de fusão. Esta técnica se mostrou atraente a ponto de inspirar também um algoritmo de seleção estática, nomeado de Seleção Baseada em Centralidade (*Centrality Based Selection - CBS*). Para a seleção, dois critérios definem o subconjunto: i) a influência do classificador medida pela posição da rede e ii) a acurácia do subconjunto formado pelos classificadores mais centrais. Um protocolo experimental robusto confirmou que as medidas de centralidade tem interessante potencial para ponderar os classificadores de um conjunto no contexto da fusão e também é um critério atraente para selecionar um conjunto de classificadores de forma gulosa. Ambos os métodos tiveram um bom desempenho em relação às respectivas abordagens da literatura. Comparado a métodos de fusão, o CBF foi superior em 73.33% dos casos mostrando enquanto que CBS foi superior em 65%.

Palavras-chave: Conjunto de classificadores, Centralidade, Diversidade, Fusão Ponderada, Seleção Estática.

Abstract

An ensemble of classifiers is employed as an alternative to a monolithic classifier to increase the performance of a Pattern Recognition System. A typical Multiple Classifier System can be divided into three phases: Generation, Selection, and Fusion. In the last phase, generally, a criterion to weight the team member's decision is used. To that, it can be used some heuristic such as precision of confidence. Then, some rule or equation is used in that criteria to obtain the score of the influence of the classifier, and it promotes competition among them in the classification process. This work presents a new ensemble fusion method, called Centrality Based Fusion, that uses the centrality concept in the context of Social Network Analysis (SNA) as a criterion to weight the vote of the classifiers. The centrality measures are applied in the context of SNA to weight the importance of each individual inside a social network, considering the relationship each one has to all others. Therefore, the idea is to derive the weight of the classifiers considering the prominence of the classifier inside the ensemble network, which reflects the pairwise relations between classifiers. We hypothesize that by the analysis of the classifiers' position in the ensemble network could be used as the score of the importance of the classifier in the fusion process. The centrality is also used to design a greedy selection approach, named Centrality Based Selection (CBS). The selection approach uses two criteria to define the subgroup of classifiers: i) the influence of the classifier measured by the position in the network and ii) the accuracy of the subgroup composed of the most central classifiers. A robust experimental protocol confirmed that the centrality measures are promising criteria to weight the classifiers vote in the fusion process and also is an interesting criterion to select a subgroup of classifiers in a greedy fashion. Both methods had a good performance compared to their respective literature approaches. Compared to other fusion approaches, CBF presented the best accuracy in 73.33% of the cases, while the CBS approach is superior in 65%.

Keywords: Ensemble of Classifiers, Ensemble Network, Centrality, Diversity, Weighted Fusion, Greedy Selection.

Lista de ilustrações

Figura 1 – Vértices mais importantes de acordo com as medidas de centralidade clássicas.	11
Figura 2 – Um Sistema de Múltiplos Classificadores composto por três fases distintas. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018).	14
Figura 3 – Funcionamento do Boosting.	18
Figura 4 – Funcionamento do Random Subspaces.	19
Figura 5 – Sistema de Múltiplos Classificadores com Seleção Estática de Classificadores.	30
Figura 6 – Sistema de Múltiplos Classificadores com Seleção Dinâmica de classificadores.	30
Figura 7 – Superprodução e Escolha (NABIHA; NADIR; MOKHTAR, 2011).	31
Figura 8 – Construção e Análise da Rede de Classificadores.	41
Figura 9 – Visão geral do método CBF.	42
Figura 10 – Teste de Friedman e teste post-hoc Nemenyi para comparando a acurácia de CBF e a literatura.	57
Figura 11 – Correlação entre a acurácia individual dos classificadores e o peso médio dos relacionamentos.	61
Figura 12 – Teste de Friedman e teste post-hoc Nemenyi comparando a acurácia de CBF numa rede simplificada e a literatura.	62
Figura 13 – Relação entre a importância (normalizada) do classificador (eixo x) e a acurácia (eixo y) para a base P2Problem com geração de pool por Bagging.	64
Figura 14 – Redes formadas por 20 classificadores gerados pelo método Bagging para a base P2Problem.	67
Figura 15 – Comparação pareada de <i>CBF: Grau de Entrada</i> N_{10} com métodos da literatura.	71
Figura 16 – Teste de Friedman e teste post-hoc Nemenyi comparando <i>CBF: Grau de Entrada</i> N_{10} com métodos da literatura.	72
Figura 17 – Distribuição das classes (I e II) para a base de dados P2Problem.	76
Figura 18 – Comparação pareada de CBS com a literatura.	82
Figura 19 – Teste de Friedman e post hoc Nemenyi comparando CBS com a Literatura.	83

Figura 20 – Comparação entre abordagens de seleção quanto a acurácia média e o tamanho médio do conjunto de classificadores gerados por Boosting. . .	84
Figura 21 – Comparação entre abordagens de seleção quanto a acurácia média e o tamanho médio do conjunto de classificadores gerados por Bagging. . .	103
Figura 22 – Teste de Friedman e post hoc Nemenyi comparando CBS com a Literatura. O <i>pool</i> foi gerado por Bagging.	104
Figura 23 – Comparação de CBS com a literatura quando o pool é gerado por Bagging.	104

Lista de tabelas

Tabela 1	–	Relação entre o par de classificadores c_i e c_j	20
Tabela 2	–	Interpretação da medida de centralidade quanto ao peso da aresta. (\downarrow) Quanto menor o valor do peso da aresta, maior é o valor da métrica. (\uparrow) Quanto maior o valor do peso da aresta, maior é o valor da métrica.	47
Tabela 3	–	Diversidade expressa como um valor alto (\uparrow) ou baixo (\downarrow) dependendo da medida de centralidade que avalia a importância dos membros da rede.	47
Tabela 4	–	Parâmetros indicados para o uso de CBF como método de fusão. (V) significa VERDADEIRO enquanto (F) significa FALSO.	48
Tabela 5	–	Características principais de cada problema de classificação. Número de Instâncias (# I). Número de atributos (# F). Número de Classes (# C). Taxa de Balanceamento (T.B.).	54
Tabela 6	–	Um caso de sucesso com a <i>Centralidade de Grau de Entrada</i> estimada em relacionamentos pareados N_{10} . A coluna (rel média) representa a média do peso dos relacionamentos considerando apenas os vizinhos que apontam para o vértice que é identificado por <i>id</i>	60
Tabela 7	–	Classificadores ordenados pela centralidade. A coluna (id) refere-se a identificação do classificador. A coluna (valor) indica o peso atribuído pela medida de centralidade e na coluna (valor N.) é representado o valor de centralidade normalizado. Os valores foram obtidos pela análise da base P2Problem. Foram gerados 20 classificadores para este exemplo.	66
Tabela 8	–	Acurácia média e o desvio-padrão de cada abordagem avaliada baseada em validação cruzada 6-fold. WS significa teste dos postos sinalizados de Wilcoxon, portanto, os valores representam o valor de p, e + é para resultados significativos. Os melhores resultados estão em negrito. # Best é o número de melhores resultados obtidos em todos os problemas de classificação.	69
Tabela 9	–	Classificadores gerados por Bagging ordenados pelo peso da <i>Centralidade de Grau de Entrada</i> em uma rede de relacionamentos N_{10} . (*) significa que o classificador pertence ao subgrupo ótimo.	77

Tabela 10 – Classificadores gerados por Boosting com validação cruzada (k-fold com $k = 6$) ordenados pelo peso da *Centralidade de Grau de Entrada* em uma rede de relacionamentos N_{10} . (*) significa que o classificador pertence ao subgrupo ótimo. 78

Tabela 11 – Comparação de CBS com métodos da literatura. Os valores apresentados são a taxa de reconhecimento obtida por k-fold com $k = 6$. O valor ao lado ($|C'|$) refere-se ao tamanho médio do subconjunto selecionado. Em negrito está o melhor resultado daquele problema de classificação. # Best é quantidade de melhores resultados atingidos por um método. WS significa teste de sinais de Wilcoxon, sendo que o valor representado é referente ao valor de p, enquanto que (+) é para sinalizar resultados significativos com $\alpha = 0.05$ 81

Tabela 12 – Comparação de CBS com métodos da literatura. O *pool* de classificadores foi gerado por Bagging. Os valores apresentados são a taxa de reconhecimento obtida por k-fold com $k = 6$. O valor ao lado ($|C'|$) refere-se ao tamanho médio do subconjunto selecionado. Em negrito está o melhor resultado daquele problema de classificação. # Best é quantidade de melhores resultados atingidas por um método. WS significa teste de sinais de Wilcoxon, sendo que o valor representado é referente ao valor de p, enquanto que (+) é para sinalizar resultados significativos com $\alpha = 0.05$ 102

Lista de abreviaturas e siglas

CR	Ciência das Redes
ARS	Análise de Redes Sociais
<i>pool</i>	Conjunto de classificadores
<i>ensemble</i>	Conjunto de classificadores
SMC	Sistemas de Múltiplos Classificadores
CBF	Centrality Based Fusion
CBS	Centrality Based Selection
AGOB	Aggregation Ordering in Bagging
POBE	Pruning in Ordered Bagging Ensembles
DREP	Diversity Regularized Ensemble Pruning
FS	Forward Search
BS	Backward Search
MV	Voto Majoritário
WMV	Voto Majoritário Ponderado
PV	Ponderação por Valor de Poder
RSWV	Re-scaled weighted vote
QBWWV	Quadratic best-worst weighted vote
BWWV	Best-worst weighted vote
KWMV	Voto Majoritário Ponderado proposto por Kuncheva
PW	Ponderação por Desempenho
BC	Ponderação por Combinação Bayesiana

Lista de símbolos

C	Conjunto de Classificadores
C'	Subconjunto de Classificadores
C_i	Classificador
x_i	Instância, exemplo ou observação
S	Conjunto de dados, observações ou instâncias
S_{train}	Conjunto de treinamento
S_{val}	Conjunto de validação
S_{test}	Conjunto de teste
ϖ	Medida de diversidade estimada
Ψ_i	Peso do classificador C_i
χ	Medida de centralidade
G	Grafo

Sumário

1	INTRODUÇÃO	1
1.1	Hipótese, Questão de Pesquisa e Motivação	2
1.2	Proposta	2
1.3	Objetivos	3
1.4	Contribuições	3
1.5	Estrutura do Trabalho	4
2	FUNDAMENTAÇÃO TEÓRICA	6
2.1	Ciência das Redes	6
2.1.1	Um breve histórico	6
2.1.2	Definições	7
2.1.3	Medidas de Centralidade	8
2.1.3.1	Centralidade de grau K_{c_i}	9
2.1.3.2	Centralidade de Intermediação B_{c_i}	9
2.1.3.3	Centralidade de Proximidade C_{c_i}	10
2.1.3.4	Centralidade de Autovetor	10
2.1.3.5	Centralidade Local (Q_{c_j})	11
2.1.3.6	Síntese das centralidades	11
2.1.4	Simplificação de Redes	12
2.2	Conjunto de Classificadores	13
2.2.1	Conceitos básicos	13
2.3	Geração de Classificadores	15
2.3.1	Bagging	16
2.3.2	Boosting	18
2.3.3	Random Subspaces	19
2.4	Medidas de Diversidade	19
2.4.1	Estatística Q (QS)	21
2.4.2	Coeficiente de Correlação (CC)	21
2.4.3	Desacordo (Dis)	21
2.4.4	Falta Dupla (DF)	22
2.4.5	Estatística Kappa (KS)	22

2.5	Considerações finais	23
3	ESTADO DA ARTE	25
3.1	Fusão Estática de Classificadores	25
3.1.1	Fusão de Rótulo de Classe Não Treinável	25
3.1.2	Fusão por Rótulo de Classe Treinável	26
3.1.2.1	Voto Majoritário Ponderado Clássico (WMV)	26
3.1.2.2	Voto Majoritário Ponderado proposto por Kuncheva (KWMV)	27
3.1.2.3	Ponderação por Desempenho (PW)	27
3.1.2.4	Combinação Bayesiana (BC)	27
3.1.2.5	Valor de Poder (PV)	28
3.1.2.6	Métodos inspirados pela autoridade de um classificador (RSWV, BWWV e QBWWV)	28
3.2	Seleção Estática de Classificadores	29
3.2.1	Estratégias Gulosas ou Gananciosas	32
3.2.1.1	Forward Search (FS)	34
3.2.1.2	Backward Search (BS)	34
3.2.1.3	Kappa Pruning	36
3.2.1.4	Seleção de Subconjunto de Classificadores Regulada por Diversidade (DREP)	36
3.2.1.5	Aggregation Ordering in Bagging (AGOB)	37
3.2.1.6	Pruning in Ordered Bagging ensembles (POBE)	38
3.3	Considerações finais	38
4	METODOLOGIA PROPOSTA	40
4.1	Rede de Classificadores	40
4.2	Fusão Estática dos Classificadores (CBF)	41
4.2.1	Produção	42
4.2.2	Relação Pareada	43
4.2.3	Combinação Ponderada	43
4.2.4	O Algoritmo CBF	44
4.3	Método de seleção estática (CBS)	48
4.3.1	O Algoritmo CBS	49
4.4	Considerações finais	52
5	EXPERIMENTOS	53
5.1	Avaliação experimental de CBF	55
5.1.1	Avaliação das medidas pareadas com as medidas de centralidade	55

5.1.1.1	Avaliação das centralidades numa rede ponderada e simplificada	61
5.1.2	Relação entre a acurácia dos classificadores e a centralidade	63
5.1.3	Comparação com métodos de fusão da literatura	68
5.1.4	Discussão acerca de CBF	71
5.2	Avaliação experimental de CBS	74
5.2.1	Comparação com Subconjunto Ótimo	75
5.2.2	Comparação de CBS com a literatura	80
5.2.3	Discussão acerca de CBS	83
6	CONCLUSÃO	87
6.1	Trabalhos Futuros	89
	REFERÊNCIAS	90
	APÊNDICES	100
	APÊNDICE A – CBS COM BAGGING	101

1 Introdução

Nas últimas décadas, a Análise de Redes Sociais (ARS) emergiu como um campo multidisciplinar pertencente à Ciência das Redes (CR) no qual pesquisadores buscam o entendimento de fenômenos e comportamentos utilizando redes (LATORA; NICOSIA; RUSSO, 2017). Diferentes tipos de pesquisa têm usado CR para descobrir novos padrões que dificilmente seriam observados por abordagens tradicionais. Uma das contribuições mais significativas da pesquisa conduzida por Moreno (MORENO, 1934) em 1934, um dos precursores do desenvolvimento da ARS, foi considerar o conjunto de pessoas de sua pesquisa como um todo e analisar como a escolha delas afetava a estrutura do grupo. Outro exemplo de ARS é o estudo recente de Jackson (JACKSON, 2017), no qual ele descreveu o padrão social das elites medievais da Escócia usando uma base de dados conhecida como PoM. Neste estudo foi então observado que um padrão usando métodos tradicionais não seria possível encontrar, mas que, segundo o autor, foi identificado pelo emprego de técnicas de ARS. O autor identificou um novo líder, uma pessoa que desenvolveu um papel de destaque naquela sociedade que possivelmente foi responsável por importantes acontecimentos na história da Escócia. Recentemente, o CR/ARS tem recebido a atenção de pesquisadores na área de Reconhecimento de Padrões para aplicações diferentes como Sistemas Multi-Agentes (ENEMBRECK; BARTHÈS, 2013), Detecção de Conceito (GOMES; ENEMBRECK, 2014; BARDDAL; GOMES; ENEMBRECK, 2014), Sistemas de Recomendação (MA; KING; LYU, 2009), e o Comitê de Classificadores (TRAWINSKI; CORDON, 2016). A crença é que, assim como Jackson, possam ser encontrados padrões por meio de técnicas de CR que outros tipos de abordagem talvez não consigam.

No Reconhecimento de Padrões, especificamente em tarefas de classificação, o tamanho da base de dados disponível para treinamento, o número de classes, a quantidade de atributos e a separabilidade de classes afetam o desempenho de qualquer classificador. Por consequência, um único classificador é geralmente incapaz de reconhecer com sucesso a grande variedade de padrões em um dado domínio de problemas. Diante deste cenário, surge como alternativa promissora aos sistemas monolíticos, os Sistemas de Múltiplos Classificadores (SMC). Um SMC é formado por três fases: Geração, Seleção e Combinação. A fase de geração é responsável por criar um grupo de classificadores suficientemente acurados e diferentes entre si. Na fase de seleção, escolhe-se a partir do grupo de classificadores inicial (*pool*) aqueles que podem aumentar ou manter a acurácia do conjunto, formando

um subgrupo. Já na fase de combinação, aplica-se alguma regra para ponderar o voto de mais de um classificador que pode ser proveniente do pool gerado ou do subconjunto selecionado. Cada uma destas fases possui desafios distintos.

Diante deste cenário, este trabalho propõe medir a importância dos classificadores baseados na posição que estes ocupam em uma rede. Para isso, um conjunto de classificadores e a estimativa de diversidade pareada entre eles compõe uma rede de classificadores, a qual é representada por um grafo. A partir desta estrutura, a importância dos classificadores é estimada por medidas de centralidade de Redes Complexas. A pontuação referente à importância do classificador pode então ser usada em esquemas de seleção e fusão. Portanto, um método de fusão e outro de seleção são apresentados. Um protocolo experimental foi realizado para validar os métodos propostos, bem como comparar seu comportamento perante abordagens presentes na literatura.

1.1 Hipótese, Questão de Pesquisa e Motivação

A hipótese deste trabalho é de que a centralidade de um classificador, estimada em uma rede formada por classificadores e seus relacionamentos, pode ser útil na determinação da importância dos membros de um conjunto de classificadores (*ensemble*) em processos de fusão ou seleção estática. A questão de pesquisa determinada neste estudo é: a importância do classificador medida pela posição que ele ocupa na rede é uma boa medida para avaliar a influência dos classificadores de um dado *ensemble*? A motivação desta pesquisa está relacionada com a avaliação dos classificadores para uso em métodos de seleção e fusão (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012; LI; YU; ZHOU, 2012; DAI; YE; LIU, 2017; AHMED et al., 2018). Pela perspectiva da Análise de Redes Sociais, acredita-se que o estudo de um SMC envolvendo a representação do *pool* como uma rede de classificadores (TRAWINSKI; CORDON, 2016) pode trazer avanços para o Reconhecimento de Padrões.

1.2 Proposta

Com o objetivo de avaliar a aplicação das medidas de centralidade em fases distintas de um SMC, propõe-se diferentes algoritmos capazes de representar um *pool* de classificadores como uma rede e, a partir desta, estimar e usar a influência dos classificadores no processo classificatório. Para a fase de fusão, o algoritmo proposto usa a influência do classificador (centralidade) como um critério de diferenciação entre os membros do conjunto, a fim de estimular a concorrência entre eles pela atribuição de classe para

uma instância não rotulada. Na fase de seleção, um algoritmo é proposto para utilizar a centralidade como um critério complementar ao critério da acurácia para a determinação do subconjunto.

1.3 Objetivos

O objetivo principal do presente trabalho consiste em avaliar o uso da centralidade, oriunda da teoria de redes complexas, como forma de ponderar ou selecionar estaticamente os classificadores de um *ensemble*. Para tal, o *pool* de classificadores é representado por vértices e seus relacionamentos pareados como arestas, que formam uma estrutura chamada de rede de classificadores.

Os objetivos específicos são:

- Analisar os conceitos relacionados a sistemas baseados em múltiplos classificadores com foco em seleção estática e fusão estática;
- Examinar métricas de centralidade de Redes Complexas e Análise de Redes Sociais;
- Identificar diferentes métricas pareadas para classificadores;
- Avaliar o impacto de diferentes métricas de centralidade usadas para ponderar um conjunto de classificadores no processo de fusão, relacionando-as com a acurácia do ensemble;
- Determinar o impacto de uma métrica de centralidade, estimada em relacionamentos pareados de assimetria, como um dos critérios para a seleção estática de classificadores inspirada em técnicas de seleção gulosa/gananciosa.

1.4 Contribuições

O estudo de diferentes Sistemas de Múltiplos Classificadores sugere que a exploração da importância dos classificadores ainda requer melhorias. Em (TRAWINSKI; CORDON, 2016), os autores usaram uma rede de classificadores para visualizar o *ensemble*, sugerindo que métodos de seleção e de fusão poderiam ser realizadas. Portanto, espera-se que pela realização desta pesquisa, os novos métodos de seleção e fusão de classificadores contribuam para o avanço da área de Reconhecimento de Padrões. As contribuições deste trabalho são apresentadas na sequência:

- A estimativa da importância dos classificadores, realizada por meio da análise feita por medidas de centralidade em relacionamentos pareados, apresenta pontuação diferente. Assim, este estudo apresenta 21 novas medidas para avaliar a influência classificadores.
- A pontuação proveniente da centralidade impacta diretamente no processo de fusão estática ponderada de classificadores. Este critério é competitivo se comparado a literatura;
- Cada medida de centralidade pontua os classificadores diferentemente. Além disso, a ordem dos classificadores também é diferente quando baseada nesta pontuação. Isto impacta diretamente no processo de seleção gananciosa que utilize a centralidade dos classificadores como critério para selecioná-los a partir de um *pool*;
- A seleção gulosa proposta utiliza um critério novo de complementariedade baseada na centralidade. Este método apresenta-se diferente da literatura e possui uma taxa de reconhecimento atrativa para um tamanho de *pool* bastante reduzido se comparado ao tamanho original.

Espera-se que tais observações e métodos possam contribuir para o avanço da área de Reconhecimento de Padrões e Aprendizado de Máquina, promovendo uma nova perspectiva inspirada pela análise da centralidade dos membros de uma rede de classificadores.

1.5 Estrutura do Trabalho

Este trabalho apresenta após a introdução, a Revisão da Literatura a qual pode trazer importantes conceitos novos ao leitor da área de Aprendizado de Máquina. A partir dos conceitos empregados em Ciência das Redes, foi estabelecido o relacionamento destes com o conjunto de classificadores. Portanto, diferentes métodos de seleção e fusão são apresentados no capítulo terceiro, contextualizando a importância da proposta. No quarto capítulo está apresentado o método proposto. Este, sugere que a seleção e combinação estática de classificadores seja feita utilizando os conceitos apresentados na seção de revisão com enfoque na análise da centralidade dos classificadores, a qual tem como entrada uma rede de classificadores. O capítulo quinto apresenta os experimentos já realizados. A descrição detalhada a respeito da condução destes testes teve por objetivo apresentar diferentes cenários, incrementais, o qual o mais básico experimento traz a importância dos classificadores estimada a partir de seus relacionamentos. A importância do classificador

serviu de base para um método de fusão e outro de seleção os quais são analisados separadamente. A seguir são apresentadas as conclusões e a sugestão de trabalhos futuros. Por fim, apresenta-se as referências bibliográficas utilizadas.

2 Fundamentação teórica

Neste capítulo são descritos os conceitos utilizados neste trabalho. Inicialmente, são apresentados os conceitos referentes a área de Ciência de Redes, que concentra as áreas de Redes Complexas e Análise de Redes Sociais. Passa-se então a apresentar os conceitos inerentes ao Conjunto de Classificadores. A partir deste esclarecimento, relaciona-se as duas áreas no próximo capítulo, o qual apresenta a metodologia proposta.

2.1 Ciência das Redes

A Ciência das Redes é um campo multidisciplinar que se propõe a análise de objetos pela interação observada entre eles. A teoria por trás desta área é continuamente melhorada por pesquisadores das mais diferentes áreas como a Análise de Redes Sociais, Redes Complexas, Química, Engenharia de Tráfego, Ciência da Computação entre outras. Nesta seção estão apresentados os conceitos referentes a Ciência das Redes que inspiraram a análise de um conjunto de classificadores no domínio do Aprendizado de Máquina e Reconhecimento de Padrões. Primeiramente, um histórico breve sobre a Ciência das Redes serve para o leitor não familiarizado com a área ter uma noção das importantes contribuições, as direções atuais e futuras. Em seguida, algumas importantes definições são apresentadas de forma a introduzir a estrutura de uma rede de objetos e como é realizada a análise da importância dos membros de um dado conjunto.

2.1.1 Um breve histórico

De acordo com pesquisadores de Redes Complexas (NEWMAN et al., 2006; CONTRACTOR; FORBUSH, 2017; ZWEIG, 2016), a história da Ciência das Redes começa em 1796 com Leonard Euler. Euler propôs a representação do problema conhecido das pontes de Königsberg como um grafo para encontrar um caminho que cruzasse todas as pontes uma única vez. Ele provou que tal caminho não existia observando o grau de cada vértice (o número de linhas indicando os vizinhos diretos de um vértice). A contribuição de Euler ajudou a desenvolver a Teoria de Grafos. Como declarado em (NEWMAN et al., 2006), a Teoria de Grafos é importante para descrever as propriedades topológicas de um problema. Em (ERDÖS; RÉNYI, 1959), nos anos 50, Erdős e Rényi propuseram um modelo que assumia que as conexões em uma rede eram geralmente criadas de forma

aleatória. Em 1998, Watts e Strogatz (WATTS; STROGATZ, 2008) observaram que redes aleatórias também apresentavam propriedades de mundo pequeno. Estas propriedades são o baixo comprimento médio do caminho e o elevado agrupamento local. Em 1999 Barabási e Albert (BARABÁSI; ALBERT, 1999) observaram que a World Wide Web (WWW), uma rede crescente, apresenta a propriedade de ligação preferencial, isto é, um novo vértice é ligado a uma rede preferencialmente por vértices que já se encontram bem conectados. Esta observação acaba na direção oposta das observações no contexto das redes aleatórias. A ARS tem estudado as propriedades de grafo desde a publicação do estudo conduzido por Moreno em 1934 (MORENO, 1934), o qual se propôs a analisar a sociabilidade de seus estudantes. Sociólogos geralmente se referem a ele como um dos precursores do campo da ARS (FREEMAN, 2004).

Muitas áreas de pesquisa tem usado as descobertas em Teoria dos Grafos, Redes Complexas e Análise de Redes Sociais, entretanto, hoje em dia essas áreas fazem parte de um campo mais abrangente, chamado de Ciência das Redes (LATORA; NICOSIA; RUSSO, 2017; LEWIS, 2009). A importância de se fundir essas áreas é porque elas observaram diferentes propriedades que estão presentes em diversas redes do mundo real. O uso de CR em artigos científicos é crescente, principalmente desde o final dos anos 90 (LATORA; NICOSIA; RUSSO, 2017). Os aspectos teóricos da CR que estão descritos nesta proposta estão apresentados na seção a seguir.

2.1.2 Definições

A teoria de CR utilizada para descrever e analisar objetos é extensa. Neste trabalho, são descritos apenas os aspectos mais importantes relacionados a pesquisa conduzida. Isto tem como objetivo ajudar o leitor a entender a contribuição inspirada por CR para o Conjunto de Classificadores. A CR geralmente representa o problema em estudo como um grafo, no qual os vértices representam objetos, grupos ou indivíduos, enquanto arestas representam o relacionamento entre eles. As arestas podem ser não-direcionadas no caso de relações simétricas, usadas para representar relações de diferença. As arestas também podem ser direcionadas. Neste caso, as relações assimétricas são representadas como a dependência entre um e outro objeto. Os grafos com somente arestas não-direcionadas são chamados de grafos não direcionados, ou são chamados de grafos direcionados na presença de arestas direcionadas. Um grafo não direcionado $G(V, E)$ é formado por dois conjuntos, $V \neq \emptyset$ e E . O conjunto $V = \{v_1, v_2, \dots, v_T\}$ representa objetos (elementos, indivíduos, grupos) que são chamados vértices ou nós. O conjunto $E = \{e_1, e_2, \dots, e_U\}$ é um grupo

que representa pares de objetos de V , portanto, cada elemento e é formado por um par de vértices i e j , logo $e(i, j)$. Estes objetos, i e j são considerados vizinhos, e uma aresta ligando diretamente estes objetos define que o vértice i é um vizinho direto de j . Para grafos não-direcionados, a relação (i, j) e (j, i) representa a mesma aresta.

Uma caminhada $walk(x, y)$ do vértice x para y é definido por Latora et al. (LATORA; NICOSIA; RUSSO, 2017) como uma alternada sequência de vértices e arestas $walk = (x \equiv v_0, e_1, v_1, e_2, \dots, e_l, n_l \equiv y)$. Começa do vértice x e termina no vértice y , sendo cada aresta representada por $e_i = (n_{i-1}, n_i)$ para $i = 1, 2, \dots, l$. O comprimento de uma caminhada $\ell(walk)$ é o número de arestas na sequência. Um caminho é uma caminhada a qual cada vértice é visitado uma única vez. O caminho mais curto, também conhecido como geodésico, é um caminho de mínimo comprimento de um vértice x para y . O termo 'distância' é frequentemente utilizado para descrever a alternância entre vértice/aresta que separam dois vértices específicos. Neste trabalho, consideramos um grafo $G(V, E)$, no qual $V = \{c_1, c_2, \dots, c_T\}$ representa o conjunto de classificadores onde cada c_i é membro de um time composto por T classificadores. A relação entre os classificadores são medidas aos pares, tal como medidas de diversidade pareadas. A partir desta estrutura chamada de rede de classificadores, é possível estimar a importância deles por meio de Medidas de Centralidade. Estas medidas estão descritas na próxima seção.

2.1.3 Medidas de Centralidade

As medidas de centralidade foram desenvolvidas para ponderar membros de uma rede baseando-se na importância topológica que eles apresentam em uma rede. Em outras palavras, estas medidas avaliam a importância dos indivíduos considerando o papel deles em relação aos demais, utilizando-se da interpretação do relacionamento representado na rede. São chamadas de clássicas as seguintes medidas: Grau (*Degree*) (FREEMAN, 1978), Intermediação (*Betweenness*) (ANTHONISSE, 1971), Proximidade (*Closeness*) (FREEMAN, 1978) e Autovetor (*Eigenvector*) (BONACICH, 1972). Estas medidas tem foco em diferentes pedaços de informação, tais como a) o número de arestas (centralidade de grau); b) geodésico (proximidade, intermediação), e c) caminhadas (eigenvector) também conhecido como refinamento iterativo em (LÜ et al., 2016).

2.1.3.1 Centralidade de grau K_{c_i}

A centralidade de grau K_{c_i} está relacionada com o número de arestas diretamente ligadas a um membro da rede. Esta medida é apresentada na Equação 2.1:

$$K_{c_i} = \sum_{c_j=1}^T E_{c_i c_j} \quad (2.1)$$

onde a aresta $E_{c_i c_j}$ conecta os membros c_i e c_j , enquanto T é o total de membros da rede (classificadores). Note que o peso associado a aresta $E_{c_i c_j}$ pode ser o peso original da aresta (Grau Ponderado) ou simplesmente 1.0 (Grau Não-Ponderado), o qual serve para indicar a presença de uma aresta. No caso do Grau Não-Ponderado, o grau máximo de um vértice é $K_{c_i} = \{T - 1\}$, isto para uma rede composta por apenas um componente. Para definir a quantidade de componentes da rede, observa-se a quantidade de sub-redes as quais cada vértice possui um caminho para qualquer outro vértice apenas pela alternância entre vértice e aresta. Se todos os vértices desta rede podem ser alcançados, então essa rede possui apenas um componente.

A *Centralidade de Grau* pode ser a mais simples de ser estimada. Ela considera apenas o aspecto local da rede para a realização da estimativa. Desta forma, o grau do vértice é definido pela quantidade de ligações diretas com seus vizinhos. Quando o peso destas arestas (ligações) é considerado, tem-se o grau ponderado do vértice. Caso o peso da aresta seja desconsiderado, torna-se impraticável a sua estimativa em redes completas, pois o grau dos classificadores é o mesmo ($T - 1$). Uma alternativa para a estimativa do grau não-ponderado é a simplificação da rede (ZHOU; MAHLER; TOIVONEN, 2012; DIANATI, 2016).

2.1.3.2 Centralidade de Intermediação B_{c_i}

Outra medida de centralidade clássica é a de intermediação (ANTHONISSE, 1971; FREEMAN, 1978). Esta medida considera o número de geodésicos de cada membro da rede para os demais que passam por um vértice em particular, como demonstrado pela Equação 2.2:

$$B_{c_i} = \sum_{c_j c_k} \frac{g_{c_j c_k}^{c_i}}{g_{c_j c_k}} \quad (2.2)$$

onde $g_{c_j c_k}^{c_i}$ é o número de geodésicos entre c_j e c_k que passam por c_i . O total de geodésicos entre c_j e c_k é $g_{c_j c_k}$. Esta medida de centralidade tem elevado custo computacional, mas

os autores em (BRANDES, 2008) sugerem uma diferente estratégia para o cálculo dos geodésicos, o que permite estimar a centralidade de intermediação a um menor custo.

2.1.3.3 Centralidade de Proximidade C_{c_i}

Uma medida comum que também usa os geodésicos é a *Centralidade de Proximidade* (FREEMAN, 1978; WASSERMAN; FAUST, 1994). Esta centralidade estima a distância média entre um membro da rede para os todos os demais, levando em consideração o comprimento médio dos geodésicos. Ela considera que membros com alta centralidade são aqueles que são próximos aos demais. Assim como a *Centralidade de Intermediação*, esta medida também depende de uma rede formada por somente um componente. A Equação 2.3 apresenta como é estimada a *Centralidade de Proximidade*:

$$C_{c_i} = \frac{1}{l_i} = \frac{|T|}{\sum_{c_j} g_{c_i, c_j}} \quad (2.3)$$

onde l_i é a média do geodésico de cada membro c_i para todos os outros membros c_j são estimadas. A média da distância, quando baixa, significa alta centralidade do membro. Para redes com mais componentes, a *Centralidade de Intermediação* e a *Centralidade de Betweenness* podem ser calculadas como indicado em (BOLDI; VIGNA, 2014).

2.1.3.4 Centralidade de Autovetor

A *Centralidade de Autovetor* é também uma medida de centralidade clássica (BONACICH, 1972; BORGATTI, 2005; BONACICH, 2007). Ela é semelhante a Centralidade de Grau a princípio, mas considera não só os vizinhos diretos de cada vértice, considera a centralidade destes vizinhos diretos em relação aos vizinhos diretos deles. Desta forma, o membro mais central é obtido se ele tem relacionamentos com vizinhos os quais também ocupam uma posição central na rede (RUHNAU, 2000). A Equação 2.4 apresenta esta centralidade de duas maneiras, conforme proposta por Bonacich (BONACICH, 1972):

$$\lambda x = Ax, \lambda x_i = \sum_{j=1}^n a_{ij} x_j, i = 1, \dots, n. \quad (2.4)$$

onde A é a matriz de adjacência, λ é uma constante (o autovalor), e x é o Autovetor. A centralidade de um membro é portanto proporcional a soma da centralidade de seus vizinhos diretos.

2.1.3.5 Centralidade Local (Q_{c_j})

Uma medida de centralidade mais recente, chamada de *Centralidade Local* foi apresentada em (CHEN et al., 2012). Para apontar o membro da rede mais central, esta medida de centralidade considera os vizinhos mais próximos e seus respectivos vizinhos mais próximos. A centralidade L_{c_i} é estimada como demonstrado pela Equação 2.6:

$$Q_{c_j} = \sum_{c_w \in \Gamma_{c_j}} N_{c_w}, \quad (2.5)$$

$$L_{c_i} = \sum_{c_j \in \Gamma_{c_i}} Q_{c_j}, \quad (2.6)$$

onde Γ_{c_j} é o conjunto de vizinhos mais próximos de c_j , os vizinhos mais próximos c_w e os vizinhos diretos de c_w é representado por N_{c_w} . Esta abordagem tem seu uso limitado a redes não-ponderadas e não-completas.

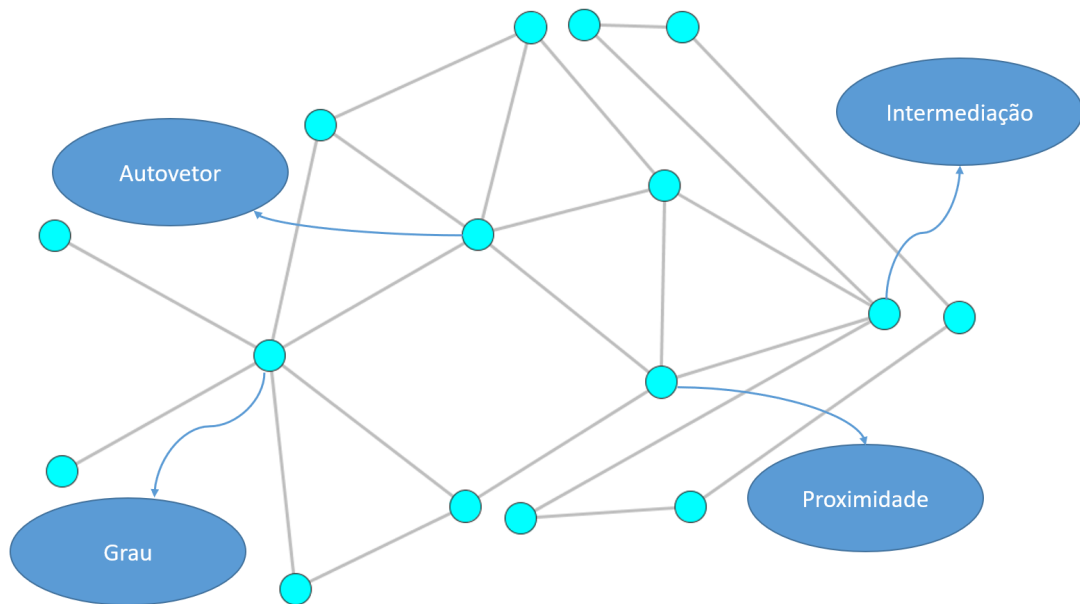


Figura 1 – Vértices mais importantes de acordo com as medidas de centralidade clássicas.

2.1.3.6 Síntese das centralidades

As medidas de centralidade descritas na Figura 1 podem ser utilizadas para redes com relacionamentos simétricos ou assimétricos. Entretanto, essas medidas são comumente usadas em relacionamentos simétricos. Para relacionamentos assimétricos, é necessário respeitar a direção das arestas. Considere as duas variantes da *Centralidade de Grau*, chamados de *Grau de Entrada* (Indegree) e *Grau de Saída* (Outdegree). A primeira medida

conta as ligações direcionadas para o vértice enquanto que a segunda é o oposto, isto é, conta a quantidade de ligações que partem de um vértice em particular para os demais.

Nesta seção, foram apresentadas algumas medidas de centralidade, como as clássicas, que também são utilizadas comumente na literatura. Cada medida foca em aspectos diferentes da rede para o entendimento do papel dos membros. Em (VALENTE et al., 2008), diversas comparações entre as medidas de centralidade são apresentadas e a conclusão é de que as medidas são diferentes apesar de estarem relacionadas entre si por um ponto de vista conceitual. A escolha da centralidade, portanto, depende do que a rede representa e o que se espera responder pela análise da posição dos membros do grupo representados na rede.

2.1.4 Simplificação de Redes

A simplificação ou poda de uma rede pode ser realizada tanto por vértice, por aresta ou ambas. Vários métodos para a remoção de aresta são apresentados em (ZHOU; MAHLER; TOIVONEN, 2012), sendo o mais simples o Algoritmo Ingênuo. Entretanto, os métodos requerem um parâmetro γ para a definição da quantidade de arestas a serem removidas. Para isso, um certo conhecimento sobre a rede é necessário. O algoritmo inicia-se ordenando as arestas e depois inicia-se a remoção de cada aresta em ordem. O algoritmo preserva a quantidade de componentes na rede, sempre evitando a remoção de uma aresta que também serve como ponte. O critério de parada é $n = \gamma \times (|E| - (|V| - 1))$, onde $|E|$ e $|V|$ é o número total de arestas e vértices, respectivamente.

Em (SILVA et al., 2018) foi proposto o Algoritmo 1 o qual exclui o parâmetro γ do Algoritmo de Simplificação Ingênuo. O critério de parada então passou a ser a identificação de uma aresta (a ser removida) que pudesse aumentar o número de componentes da rede.

O Algoritmo 1 difere do original (ZHOU; MAHLER; TOIVONEN, 2012) em dois aspectos principais. Primeiro, o critério de parada é determinado pela aresta ponte, então se ela é a próxima a ser removida o algoritmo encerra. No algoritmo original esta aresta ponte somente é ignorada, continuando o processo de poda até que o valor em γ seja atingido. O outro aspecto é de exigir a determinação do valor de γ pode exigir algum conhecimento por parte do usuário em relação ao que ele espera da análise da rede. Ambos os métodos tem por função diminuir somente os relacionamentos de uma rede enquanto a mantém como um único componente.

Algoritmo 1: O Algoritmo de Simplificação Ingênua modificado

input : Uma rede ponderada $G = (V, E)$ tal que $E = \{e_1, e_2, \dots, e_N\}$ e $V = \{c_1, c_2, \dots, c_T\}$
output : Uma sub-rede $H \subset G$ tal que $H = (V, F)$ e $F \subset E$

- 1 $F \leftarrow E$;
- 2 *OrdenarArestas*(F);
- 3 $i \leftarrow 1$;
- 4 **enquanto** $i \leq N$ **faça**
- 5 **se** $C(c_r, c_s; F \setminus e_i) \neq -\infty$ **então**
- 6 $F \leftarrow F \setminus e_i$;
- 7 **em outro caso**
- 8 **retorne** $H = (V, F)$;
- 9 **fim**
- 10 $i \leftarrow i + 1$;
- 11 **fim**

2.2 Conjunto de Classificadores

No Aprendizado de Máquina, especificamente o Aprendizado por Conjunto de Classificadores (Ensemble Learning) preocupa-se em explorar a diversidade dos membros de um conjunto para que a taxa de reconhecimento do conjunto seja melhorada em problemas de classificação. Vários modelos são combinados e a decisão final do conjunto, a predição, é determinada por ao menos um membro deste conjunto. Para métodos de combinação, os membros do conjunto influenciam a predição diferentemente. O peso dos classificadores é determinado por uma equação a qual recebe geralmente a taxa de reconhecimento individual dos membros, que pode ser obtida por uma base de validação ou de treino. Este capítulo apresenta a teoria necessária para o entendimento de como o Aprendizado de Máquina vem resolvendo problemas de classificação, com foco em métodos de combinação e seleção.

2.2.1 Conceitos básicos

O conjunto de classificadores vem sendo empregado como uma alternativa interessante ao risco de se desenvolver apenas um único classificador para rotular padrões desconhecidos em problemas de classificação. O conjunto divide a responsabilidade de cobrir todo o espaço do problema entre os membros de um time composto por classificadores com duas características fundamentais: a divergência e a acurácia. Assim, combinando decisões de classificadores que apresentam diferentes erros, é possível que o conjunto possa apresentar uma melhor taxa de reconhecimento se comparado a sistemas monolíticos, isto é,

aqueles sistemas de classificação que um único especialista computacional atribui classe a um padrão não rotulado. O *Ensemble Learning* é usado em diferentes aplicações como em fluxo de dados (data stream) (KRAWCZYK et al., 2017; GOMES et al., 2017), mudança de conceito (concept-drift) (WANG; MINKU; YAO, 2018), desbalanceamento de classes (class imbalance) (HAIXIANG et al., 2017) e análise de sentimentos (sentiment analysis) (ALDOGAN; YASLAN, 2016), apenas citando alguns.

O desenvolvimento de um conjunto de classificadores é geralmente descrito como um processo composto de três fases distintas: geração, seleção e fusão (CRUZ; SABOURIN; CAVALCANTI, 2018; Britto Jr.; SABOURIN; OLIVEIRA, 2014; BURDUK, 2017). Na primeira fase, métodos de geração proveem um conjunto de classificadores, preferencialmente diversos entre si e individualmente acurados. Já a segunda fase é responsável por encontrar um sub-conjunto de classificadores os quais podem ser mais acurados que utilizar o conjunto inteiro de classificadores. Este modelo também apresenta a vantagem de economia de recursos, pois toma decisões consultando menos especialistas. Esta fase não é obrigatória. Finalmente, tem-se a fase de fusão, a qual combina mais de um classificador por meio de alguma regra. A expectativa é que a decisão de alguns classificadores seja mais confiável que a decisão de qualquer membro do conjunto individualmente. No caso de somente um especialista fazer parte do subconjunto de classificadores (escolhido pela fase de seleção) a combinação de classificadores não é empregada.

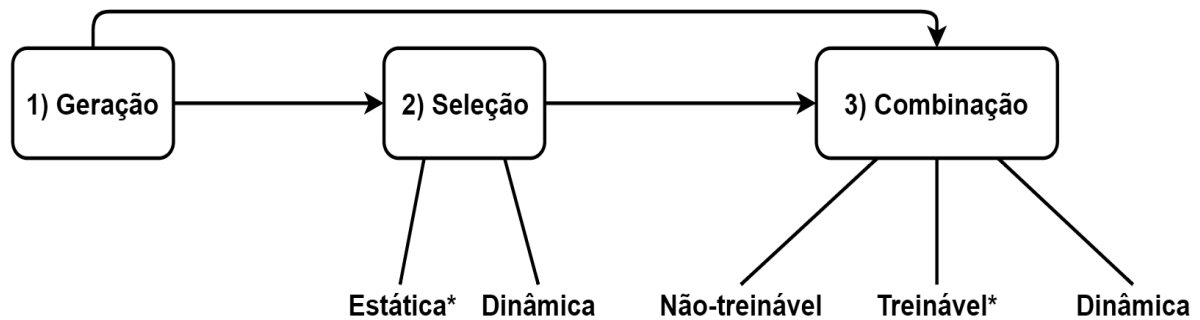


Figura 2 – Um Sistema de Múltiplos Classificadores composto por três fases distintas. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018).

Formalmente, um conjunto de classificadores é apresentado como um conjunto $C = \{c_1, c_2, \dots, c_T\}$. Cada membro deste conjunto representa uma função independente $c_t : R^n \rightarrow W$ que assinala o rótulo de uma classe $w_i \in W$ para $x \in R^n$, onde $W = \{w_1, w_2, \dots, w_M\}$. A decisão final considera as decisões de todos os classificadores ou de uma parte deles, caso sejam empregadas estratégias de seleção estática ou dinâmica (EK et al., 2015; Britto Jr.; SABOURIN; OLIVEIRA, 2014). Se o grupo final é $T \geq 2$, então um

método de fusão precisa combinar a decisão do conjunto. Esses cenários podem começar a ser notados na Figura 2 pela divisão de um Sistema de Múltiplos classificadores em partes distintas, destacando-se as sub-divisões da fase de seleção e combinação a qual esta proposta está focada.

A literatura é repleta de métodos de fusão (KITTLER et al., 1998; Ponti Jr., 2011; KUNCHEVA, 2014), sendo que é possível assumir que todos os classificadores tem desempenho semelhante e, portanto, o sistema combina as decisões do conjunto sem diferenciá-los. Considerando que os classificadores podem ser diferentes entre si, uma alternativa tem sido ponderar o voto dos classificadores enquanto assume-se que eles competem entre si para assinalar o rótulo correto de um novo caso avaliado (KUNCHEVA et al., 2003). Esta competição, promovida por pesos que estão associados a cada classificador, tem se mostrado promissora. A literatura testemunha uma ampla variedade de estratégias para a ponderação e combinação do peso dos classificadores na decisão final que dividem-se em estáticas ou dinâmicas. A alternativa estática considera que os pesos devem ser associados na fase de treinamento de um sistema de classificação, sendo que os pesos permanecem inalterados na fase de teste. Entretanto, quando os pesos são associados por estratégia dinâmica, cada classificador recebe um peso diferente para cada instância testada, ou seja, a abordagem reage a informações de cada instância não rotulada.

Diversos autores provaram que a acurácia de um conjunto depende da acurácia individual dos membros do conjunto e também da necessidade de que suas decisões precisam divergir, no sentido de que se os erros cometidos forem diferentes há uma complementariedade quanto suas expertises (LI et al., 2012; GIACINTO; ROLI, 2001; ULAS et al., 2016; KUNCHEVA; WHITAKER, 2003; Britto Jr.; SABOURIN; OLIVEIRA, 2014). Métodos de seleção fazem uso destas informações (BANFIELD et al., 2005; YIN et al., 2014; GARGIULO; MAZZARIELLO; SANSONE, 2013) enquanto que isso tem sido ignorado na fase de fusão, como o número de trabalhos na seção de trabalhos relacionados sugere. Para estimar a diversidade dos classificadores, a literatura apresenta algumas opções as quais estão descritas na próxima seção.

2.3 Geração de Classificadores

Um sistema de múltiplos classificadores é mais confiável que sistemas monolíticos (KUNCHEVA, 2014). No caso de alguns classificadores estarem incorretos, outros classificadores são capazes de fornecer uma resposta correta e manter a acurácia do conjunto

alta (VALENTINI; MASULLI, 2002). Para isso, um conjunto de classificadores precisa ser acurado e seus membros precisam ser diferentes entre si. Estes dois aspectos podem ser conflitantes. Considere dois classificadores, um que está sempre correto enquanto que o outro está sempre incorreto. Neste caso, a forte diversidade presente no conjunto não ajuda a ultrapassar a taxa de reconhecimento do melhor classificador devido a falta de precisão de um dos membros. No outro extremo tem-se classificadores idênticos os quais também não ajudam o conjunto a obter uma resposta mais precisa pela falta de complementariedade de opiniões e, além disso, ocupa mais recursos desnecessariamente.

Para a geração de um conjunto de classificadores (*pool*) há na literatura estratégias de geração heterogêneas e homogêneas. A primeira estratégia gera diferentes classificadores numa base de treinamento comum para obter opiniões diversificadas entre os especialistas (classificadores). A segunda usa o mesmo classificador base, mas varia a base de treinamento utilizada, assim, especializado em observações diferentes cada classificador apresenta opiniões divergentes de seus pares. As abordagens clássicas Bagging (BREIMAN, 1994), Boosting (SCHAPIRE et al., 1998) ou Random Subspaces (HO, 1998) são comumente utilizadas para compor um time de classificadores homogêneos. Uma revisão destes métodos pode ser encontrada em (KOTSIANTIS, 2014).

2.3.1 Bagging

Bagging (bootstrap aggregating) foi apresentado em (BREIMAN, 1994) como um método pioneiro de criação de um pool de classificadores $C = \{c_1, c_2, \dots, c_T\}$ utilizando diferentes amostras do conjunto de treinamento original $S_{train} = \{s_1, s_2, \dots, s_T\}$. Segundo o Algoritmo 2, as instâncias do conjunto de treinamento S_{train} podem aparecer em mais de uma amostra s_t (reposição) ou podem não aparecer em nenhuma das amostras (linha 2). Um classificador c_t é obtido pela indução de algum algoritmo (classificador base) em cada amostra s_t (linha 3). O conjunto de classificadores é, portanto, formado por classificadores $c_t \in C$ treinados em sua respectiva porção de treinamento $s_t \in S_{val}$. Este método tem a vantagem de poder criar todos os classificadores em paralelo devido a independência entre os classificadores e a aleatoriedade do sorteio das amostras para a composição das porções de treinamento s_t . As amostras s_t obtidas do conjunto de treinamento podem ser do mesmo tamanho ($BS = 100\%$) ou podem variar, a partir da especificação do tamanho que cada amostra deve ter.

Em (MARTÍNEZ-MUÑOZ; SUÁREZ, 2010) foi apresentado um conjunto de observações acerca do tamanho da amostra s_t definido por BS . Uma das contribuições

Algoritmo 2: Bagging (S_{val} , T , LA , BS)

Entrada: Base de validação (S_{val})
Entrada: Número de classificadores (T)
Entrada: Algoritmo de aprendizagem (classificador base) (LA)
Entrada: Tamanho de cada amostra s_i obtida de S_{val} (BS)

- 1 **para** $t \leftarrow 1$ até T **faça**
- 2 $s_t \leftarrow$ amostra *bootstrap* de S_{val} de tamanho BS ;
- 3 $c_t \leftarrow$ LA (s_t);
- 4 **fim**
- 5 **retorne** C ;

foi avaliar a quantidade de observações diferentes (dif_{obs}) que uma amostra apresenta em relação ao conjunto original da qual foi extraída. Os autores estimaram essa diferença por meio da Equação 2.7:

$$dif_{obs} \approx 1 - 1 - \frac{m}{n} \quad (2.7)$$

onde m representa o tamanho da amostra de treinamento s_t enquanto n refere-se ao tamanho da fonte desta amostra, possivelmente o conjunto de treinamento S_{train} . Por exemplo, caso $m = n$ ($\%bag = 1.0$) tem-se que apenas 63.2% das instâncias de s_t são exclusivas enquanto o restante é preenchido por instâncias repetidas. Para $m = \frac{n}{2}$ ($\%bag = 0.5$) a proporção de instâncias diferentes na amostra é de 39.3% enquanto para $m = n + \frac{n}{2}$ é 77.7%. Isto sugere que quando a diversidade entre as instâncias de uma amostra é baixa o classificador treinado tende a aprender melhor o rótulo representado pela maior parte das instâncias. Isso contribui para a diversidade do conjunto no sentido de que cada amostra usada para treinamento terá suas próprias instâncias repetidas, gerando modelos possivelmente complementares em relação ao todo.

Diante deste cenário, há um risco de que cada classificador seja bastante parecido, por isso é importante a escolha de um classificador fraco, isto é, um classificador base capaz de traçar sua hipótese reagindo a pequenas mudanças entre as amostras que diferenciam cada classificador do conjunto. Não há um tamanho padrão que se mostrou favorável ao aprendizado de uma variedade de problemas de classificação. Ainda neste estudo foi percebido que na prática um modelo cuja amostra tem de 60% a 80% do tamanho do conjunto de treinamento produz conjuntos de classificadores mais acurados que aqueles com 100%.

2.3.2 Boosting

Outra abordagem clássica para a geração de classificadores é o Boosting. Esta alternativa ao Bagging não pode gerar o conjunto de classificadores em paralelo, pois a partir de um primeiro classificador gerado aleatoriamente, cada novo classificador é formado baseando-se preferencialmente nas instâncias mais difíceis de serem aprendidas pelos modelos precedentes. A Figura 3 apresenta o funcionamento deste método de criação. A partir de um conjunto de treinamento S_{val} tem-se inicialmente as instâncias ponderadas igualmente. A partir do sorteio para a composição do primeiro classificador, todas as instâncias possuem igual chance de serem sorteadas. Após a criação do primeiro classificador, este avalia as instâncias do conjunto incrementando o peso das instâncias cujo rótulo ele é incapaz de acertar. Portanto, o segundo classificador sorteará as instâncias levando em consideração que instâncias não classificadas corretamente possuem vantagem no sorteio. Desta forma, a premissa de Boosting é reforçar padrões mais difíceis de serem classificados. Este processo se repete até que o último classificador seja criado. É importante mencionar que o método Boosting apresentado nesta seção não tem implementado o sistema de pesos comumente utilizado na fusão ponderada dos classificadores como descrito originalmente. Portanto, o Boosting é apresentado apenas como um método que na composição dos classificadores prefere as instâncias mais difíceis de serem classificadas, de forma distinta ao Bagging, que escolhe as instâncias em um processo aleatório.

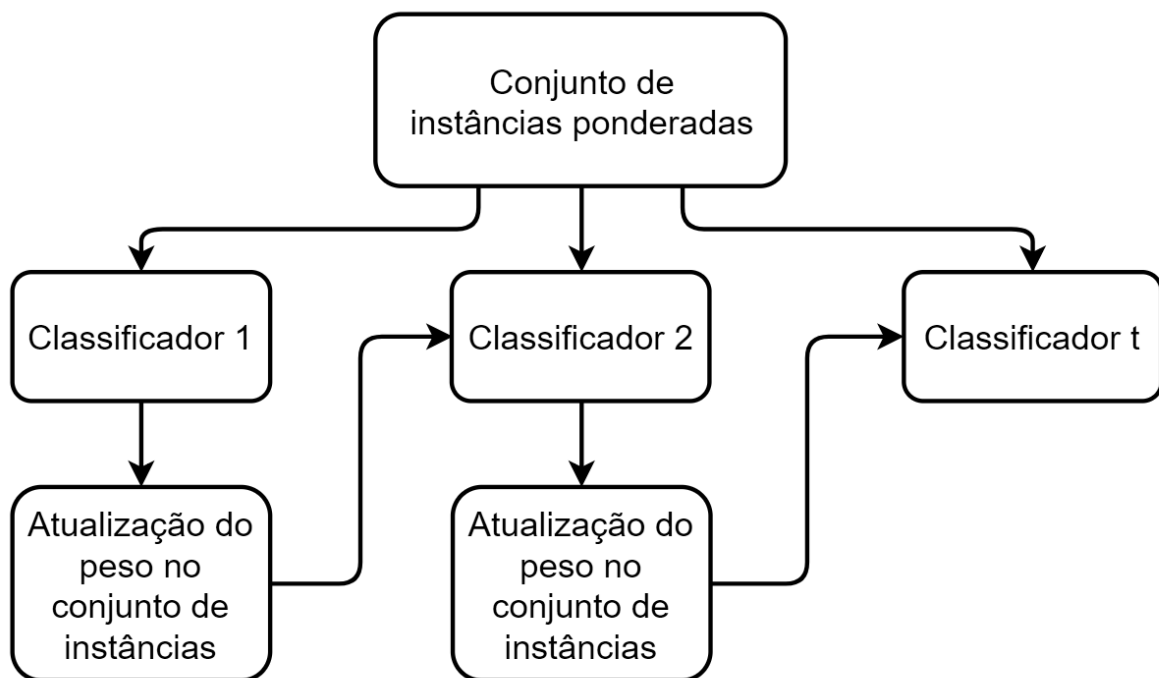


Figura 3 – Funcionamento do Boosting.

2.3.3 Random Subspaces

O método de geração Sub-Espaços Aleatórios (*Random Subspaces*) (HO, 1998) usa uma parte dos atributos (sorteados aleatoriamente) para o treinamento de diferentes classificadores. Os atributos podem estar presentes em mais de um subconjunto, entretanto, cada atributo deve aparecer apenas uma vez no conjunto que será usado para treinar cada classificador individualmente. Como um problema de classificação pode apresentar um número alto de atributos com a possibilidade de muitos deles serem redundantes, a seleção de alguns atributos facilita o aprendizado por parte do classificador base escolhido e ainda evita a maldição da dimensionalidade. A Figura 4 apresenta o funcionamento desta estratégia para a geração de um pool.

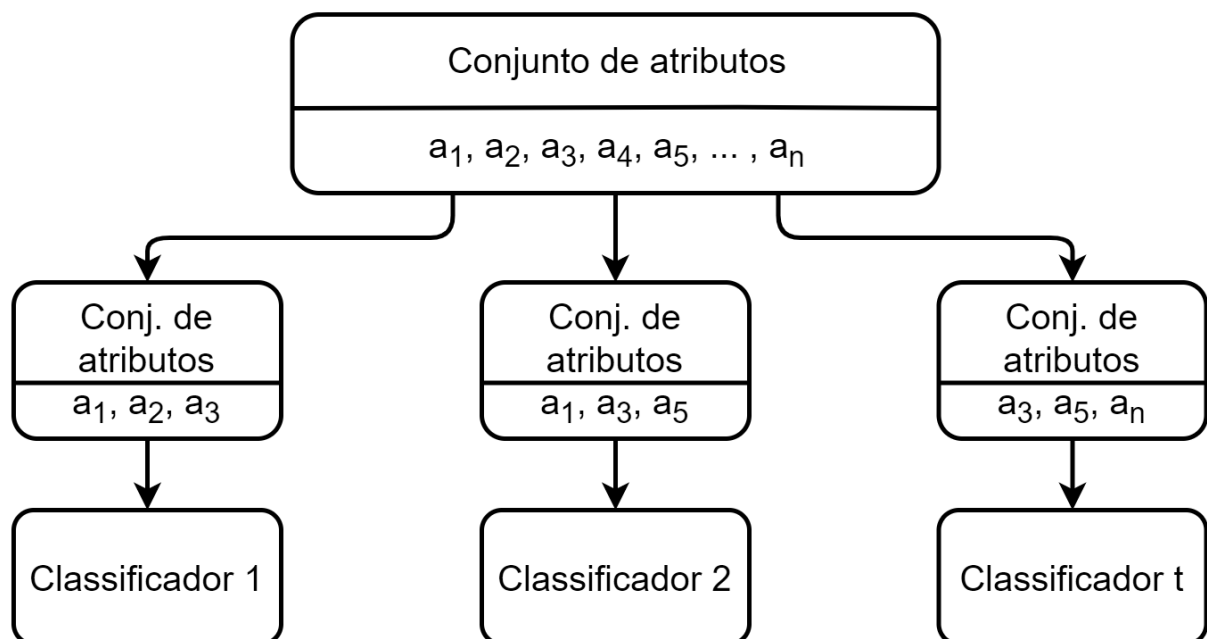


Figura 4 – Funcionamento do Random Subspaces.

2.4 Medidas de Diversidade

A diversidade entre os classificadores pode ser explorada em todas as fases de um conjunto de classificadores (geração, seleção e fusão). Para estimar a diversidade, medidas foram criadas para avaliar os classificadores aos pares, chamadas de medidas de comparação pareada (*Pairwise Measures*). Outro tipo de medidas foram criadas para avaliar o conjunto de classificadores como um todo, sendo identificadas como medidas de comparações não-pareadas (*Nonpairwise Measures*).

O primeiro conjunto de medidas, as pareadas, têm como representantes (KUNCHEVA; WHITAKER, 2003): a Estatística Q (*Q statistics*) (YULE, 1990), o Coeficiente de Correlação (*Correlation Coefficient*) (SEATH; SOKAL, 1973), o desacordo (*Disagreement measure*) (SKALAK, 1996; HO, 1998), a Falta Dupla (*Double Fault*) (GIACINTO, 2001), e a Estatística Kappa (*Kappa Statistic*) (MARGINEANTU; DIETTERICH, 1997). Os representantes do segundo grupo de medidas podem ser: a Variância de Kohavi-Wolpert (*Kohavi-Wolpert variance*) (KOHAVI; WOLPERT, 1996), o Acordo entre Avaliadores (*Interrater Agreement*) (DIETTERICH, 2000; FLEISS; LEVIN; PAIK, 2003), a Entropia (*Entropy*) (CUNNINGHAM; CARNEY, 2000), a Medida de Dificuldade (*Measure of difficulty*) (HANSEN; SALAMON, 1990), a Diversidade Geral (*Generalized diversity*) (PARTRIDGE; KRZANOWSKI, 1997) e a Falha Coincidente (*Coincident failure diversity*) (PARTRIDGE; KRZANOWSKI, 1997), apenas para citar algumas. O foco desta seção é apresentar as medidas de diversidade pareadas, pois este tipo de medida é apropriada para representar uma rede de classificadores e a partir dela podem ser feitas análises de rede, inspiradas na literatura de Análise de Redes Sociais, de Redes Complexas ou do agregador Ciência das Redes.

As medidas de diversidade pareadas são baseadas na relação predições corretas/incorretas entre dois classificadores. Tal relação entre os classificadores c_i e c_j , estão apresentadas na Tabela 1. Para uma dada instância, se ambos os classificadores estiverem corretos, então N_{11} é incrementado. Se ambos os classificadores estão incorretos, então N_{00} é incrementado. Se c_i está correto, mas c_j não está, então N_{10} é incrementado. Caso contrário, N_{01} é incrementado. Para estimar estas relações, geralmente uma base de treino ou de validação são utilizadas, evitando o *overfitting*, i.e., a análise dos classificadores pela mesma base de dados usada para a criação deles.

Tabela 1 – Relação entre o par de classificadores c_i e c_j

	c_j correto (1)	c_j incorreto (0)
c_i correto (1)	N_{11}	N_{10}
c_i incorreto (0)	N_{01}	N_{00}
<i>Total, $N = N_{00} + N_{01} + N_{10} + N_{11}$</i>		

As medidas de diversidade seguintes são baseadas nas possíveis relações descritas na Tabela 1. Note que a relação entre classificadores pode ser simétrica em N_{00} e N_{11} ou assimétrica no caso de N_{01} e N_{10} . Relações simétricas entre classificadores também são observadas em medidas de diversidade, portanto, estas medidas também podem ser utilizadas para representar a aresta entre dois classificadores apresentados como uma rede

de classificadores. Assim, tanto as relações N_{00} e N_{11} , quanto às medidas de diversidade aqui apresentadas podem ser utilizadas em uma rede não-direcionada.

2.4.1 Estatística Q (QS)

Uma medida de diversidade pareada bastante conhecida é a *Estatística Q* (QS), apresentada na Equação 2.8. Ela assume valores no intervalo $[-1,1]$. Para classificadores estatisticamente independentes, QS tem valor 0. Se os classificadores frequentemente reconhecem os mesmos objetos corretamente, QS será positivo. Caso contrário, se o par de classificadores geralmente cometem erros em diferentes instâncias de teste, então, QS será negativo.

$$QS = \frac{N_{11} \times N_{00} - N_{01} \times N_{10}}{N_{11} \times N_{00} + N_{01} \times N_{10}} \quad (2.8)$$

2.4.2 Coeficiente de Correlação (CC)

Uma medida similar a QS é o Coeficiente de Correlação (CC). Esta pode ser estimada como mostra a Equação 2.9. Para CC , os valores podem estar entre $[-1,1]$, contudo, se o valor de acurácia individual for fixo para um par de classificadores, este pode assumir outro intervalo como apresentado em (KUNCHEVA; WHITAKER, 2001). O mesmo pode acontecer com as medidas Desacordo e Falta Dupla.

$$CC = \frac{N_{11} \times N_{00} - N_{01} \times N_{10}}{\sqrt{\Delta}} \quad (2.9)$$

onde $\Delta = (N_{11} + N_{10}) \times (N_{11} + N_{01}) \times (N_{01} + N_{00}) \times (N_{10} + N_{00})$.

2.4.3 Desacordo (Dis)

A medida de Desacordo (Dis) estima o número de observações na qual um dos classificadores está incorreto enquanto o outro está correto. A diversidade é representada por uma pontuação alta no intervalo $[0,1]$. Esta medida é descrita na Equação 2.10:

$$Dis = \frac{N_{01} + N_{10}}{N} \quad (2.10)$$

2.4.4 Falta Dupla (DF)

Outra medida pareada é a Falta Dupla (DF) (GIACINTO, 2001). Ela é definida como o número de exemplos que tem sido classificados erroneamente por ambos os classificadores c_i e c_j , como descrito na Equação 2.11. Em oposição a medida Dis , DF representa a diversidade por uma pontuação baixa no intervalo $[0, 1]$.

$$DF = \frac{N_{00}}{N} \quad (2.11)$$

2.4.5 Estatística Kappa (KS)

A Estatística Kappa (KS), proposta por (MARGINEANTU; DIETTERICH, 1997), considera 3 cenários distintos: os classificadores concordam entre si para todas as instâncias avaliadas ($KS = 1$), se não há correlação entre as decisões ($KS = 0$) e $KS < 0$ se eles concordam menos que o esperado ao acaso (isto raramente acontece). Esta medida é apresentada pela Equação 2.12:

$$KS = \frac{\Theta_1 - \Theta_2}{1 - \Theta_2} \quad (2.12)$$

onde

$$\Theta_1 = \frac{N_{11} + N_{00}}{N}, \quad (2.13)$$

e

$$\Theta_2 = \frac{(N_{11} + N_{10}) \times (N_{01} + N_{00}) + (N_{11} + N_{01}) \times (N_{10} + N_{00})}{N^2} \quad (2.14)$$

As medidas apresentadas nesta seção são comumente utilizadas em um contexto estático, ou seja, para a estimativa da diversidade entre os classificadores usa-se uma base de treino S_{train} ou de validação S_{val} . Para estimar a diversidade entre os classificadores em um ambiente dinâmico, o leitor pode encontrar alternativas em (DAI; YE; LIU, 2017), (PARTALAS; TSOUMAKAS; VLAHAVAS, 2010) e em (TSYMBAL; PUURONEN; PATTERSON, 2003). Estes autores propuseram diferentes medidas para diferenciar os classificadores considerando cada novo objeto a ser rotulado, obtido numa base de teste S_{test} .

2.5 Considerações finais

Neste capítulo, foram apresentados alguns conceitos de Redes Complexas e dos Sistemas de Múltiplos Classificadores. No tópico de redes, o objetivo foi mostrar como elas representam objetos e seus relacionamentos. A rede pode ser representada por meio de um grafo, que posteriormente é analisado com técnicas como a de estimativa de centralidade, a qual aponta a importância de cada um dos objetos.

Para estimar a importância dos objetos de uma rede, há de se estabelecer um contexto, o qual é expresso na rede por meio de relacionamentos pareados (representados por arestas) entre os objetos (representados por vértices). Estes relacionamentos podem ser direcionados ou não direcionados. Caso a rede (representada por um grafo) seja direcionada, as medidas de centralidade devem respeitar esse direcionamento, pois, caso contrário, induz a uma análise errônea da importância dos objetos na rede. Outro aspecto considerado é como a centralidade prioriza os relacionamentos, isto é, para algumas centralidades o peso do relacionamento é mais importante quando é alto enquanto outras é o contrário. Neste sentido, é fundamental que a representação dos relacionamentos da rede esteja em conformidade com a expectativa da medida de centralidade que irá analisá-la.

Ao considerar as medidas de diversidade pareadas, observa-se que algumas expressam a diversidade com pesos altos e outras não. Se o propósito da medida de centralidade é avaliar a importância do classificador para a diversidade do grupo, ambas as medidas (diversidade e centralidade) devem ser consideradas na análise para evitar erros de interpretação. No caso da medida de diversidade Estatística Q, que expressa a diversidade com valores baixos, a análise pela medida de *Centralidade de Intermediação* pode ser feita diretamente, enquanto que para a *Centralidade de Grau* os relacionamentos precisam ser adaptados. Isto acontece porque a primeira medida de centralidade considera os pesos baixos dos relacionamentos como mais importantes, enquanto que a segunda é o contrário. A escolha por uma medida de diversidade ou de centralidade ainda não tem correlação com a acurácia de um *ensemble*, a qual esta pesquisa se propõe a investigar.

Várias medidas de diversidade foram apresentadas para diferenciar os classificadores de um conjunto, podendo ser do tipo não-pareada ou pareada. A primeira considera todos os classificadores do conjunto para sua estimativa, enquanto que a segunda realiza a estimativa sempre aos pares. O objetivo desta última, em geral, é avaliar se um par de classificadores é similar, diverso ou até mesmo neutro. Isto é feito pela observação de um conjunto de dados, geralmente de validação e, portanto, se os classificadores acertam ou

erram as mesmas instâncias eles são similares, enquanto que se eles divergem entre si, então são considerados diferentes. Se não é detectada a relação entre a decisão de um classificador e outro, eles são ditos neutros. Neste contexto, ainda não se tem clara a correlação de uma medida de diversidade com a acurácia de um SMC. Pesquisas tem apontado que há de se considerar o dilema da acurácia vs. diversidade na composição de um SMC robusto. Porém, a estimativa da diversidade fica limitada ao conjunto ou aos pares, pois a diversidade requer comparação. Esta barreira pode ser ultrapassada pelas medidas de centralidade, pois a análise dos relacionamentos de diversidade, podem estimar a contribuição ou a importância de cada membro de um *pool* para diversidade do grupo. Assim, no processo de construção de um SMC pode ser considerada a acurácia individual e a importância do classificador para a diversidade do grupo.

No próximo capítulo, são apresentadas abordagens encontradas na literatura de fusão e de seleção de classificadores. Ambas necessitam de um conjunto de classificadores criado por uma das técnicas de geração de *pool* apresentadas.

3 Estado da Arte

Este capítulo apresenta a literatura correspondente a duas fases distintas de um SMC: fusão e seleção. Ambas as fases são abordadas apenas no contexto estático. Portanto, tanto o peso dos classificadores para fusão, quanto o subconjunto de classificadores permanecem inalterados no processo de classificação de instâncias não rotuladas.

3.1 Fusão Estática de Classificadores

Vários trabalhos como (JAIN; DUIN; MAO, 2000), (ROKACH, 2010b), (ZHOU, 2012) e (WOŹNIAK; GRAÑA; CORCHADO, 2014) apresentam revisões de literatura a respeito de tópicos relacionados a conjunto de classificadores e, uma observação comum é que métodos de fusão ponderados são estratégias promissoras. Outro relato é que os métodos de fusão diferem-se entre si. Em (WOŹNIAK; GRAÑA; CORCHADO, 2014), os autores descrevem uma taxonomia dividida em 3 tipos: Fusão de Rótulo de Classe, Fusão Baseada em Treinamento e Fusão Baseada em Funções de Suporte. Cada tipo é baseado em a) Rótulo de Classe or b) Função de Suporte, e podem demandar treino extra (Treinável) ou não (Não Treinável). Destes, os trabalhos relacionados tem foco na Fusão de Classe Treinável. Há trabalhos apresentando exemplos a respeito de Funções de Suporte em (SINGH; SARKAR; NASIPURI, 2017; KRAWCZYK; WOŹNIAK, 2016).

3.1.1 Fusão de Rótulo de Classe Não Treinável

A Fusão de Rótulo de Classe não Treinável explora variações do Voto Majoritário (MV) (KUNCHEVA, 2004). Há 3 variações de MV (WOZNIK, 2014): Voto por Unanimidade, Maioria Simples, e Voto pela Maioria. O primeiro assinala a decisão do conjunto de classificadores se todos os membros apontam para o mesmo rótulo de classe. A Maioria Simples é quando há o acordo entre mais de 50% dos membros pelo mesmo rótulo (para problemas que apresentam apenas duas classes). O último, simplesmente assinala o voto do conjunto pela maioria dos votos num mesmo rótulo. Geralmente os pesquisadores referem-se a este último tipo quando relatam que usaram a combinação dos classificadores por meio de MV. O Voto da Maioria, ou Voto Majoritário, assume que cada classificador tem a mesma influência na decisão final, ignorando a habilidade em predizer amostras corretamente.

3.1.2 Fusão por Rótulo de Classe Treinável

Como os classificadores são em geral diferentes, considera-se que a Fusão por Rótulo de Classe (também conhecida por Votação Ponderada) podem assumir que os pesos estão entre estas categorias (KUNCHEVA, 2004; WOZNIAK, 2014):

- os pesos Ψ_{c_t} são assinalados para cada classificador;
- os pesos $\Psi_{c_t,w}$ são assinalados para cada classificador e para cada classe;
- os pesos $\Psi_{c_t}(x)$ são assinalados para os classificadores e são dependentes do espaço de características;
- os pesos $\Psi_{c_t,w}(x)$ são assinalados aos classificadores, classes e também dependem do espaço de características.

Nesta seção são apresentados apenas os métodos referentes ao primeiro grupo, isto é, abordagens que estão preocupadas em assinalar o peso Ψ_{c_t} para cada classificador, independente de classe ou espaço de características e, portanto, mais genérico. Para a ponderação de classificadores, geralmente usa-se variações do Voto Majoritário Ponderado (WMV) (KUNCHEVA, 2004), também conhecido como Voto Ponderado Simples em (MORENO-SECO et al., 2006).

3.1.2.1 Voto Majoritário Ponderado Clássico (WMV)

No WMV clássico, a decisão do conjunto é estimada considerando cada voto e também a influência de cada classificador, como descrito na Equação 3.1:

$$vote_{c_t} = \begin{cases} \Psi_{c_t} & \text{se } c_t \text{ escolhe a classe } w_i; \\ 0 & \text{caso contrário.} \end{cases} \quad (3.1)$$

onde Ψ_{c_t} é o peso definido pelo classificador c_t , e $vote_{c_t}$ representa o voto do classificador para uma classe específica w_i . Os votos são separados por classe como descreve a Equação 3.2, onde Θ_c é o número total de votos que cada classe recebe. Finalmente, a classe com a maior pontuação é considerada como a decisão final do conjunto.

$$\Theta_c = \sum_{t=1}^T vote_{c_t} \quad (3.2)$$

3.1.2.2 Voto Majoritário Ponderado proposto por Kuncheva (KWMV)

O Voto Majoritário proposto por Kuncheva (KWMV) (KUNCHEVA, 2014) estima o peso de cada classificador como definido pela Equação 3.3:

$$\Psi_{c_t} = \log \left(\frac{\alpha_{c_t}}{1 - \alpha_{c_t}} \right) \quad (3.3)$$

onde o peso Ψ de um classificador c_t é o logaritmo da divisão na qual a acurácia α do classificador c_t é o dividendo, e o divisor é o erro $(1 - \alpha_{c_t})$ do mesmo classificador. A acurácia α de cada classificador é estimada utilizando-se de uma base de validação S_{val} ou de treino S_{train} .

3.1.2.3 Ponderação por Desempenho (PW)

A Ponderação por Desempenho (Performance Weighting - PW) (ROKACH, 2010b) normaliza a acurácia do classificador para utilizar isso como a influência do classificador no conjunto, considerando isso como a proporção da acurácia do classificador perante o somatório de todos os membros, como demonstrado pela Equação 3.4:

$$\Psi_{c_t} = \frac{1 - \epsilon_{c_t}}{\sum_{j=1}^T \epsilon_{c_j}} \quad (3.4)$$

onde ϵ_{c_t} é o erro normalizado do classificador estimado possivelmente por uma base de validação S_{val} .

3.1.2.4 Combinação Bayesiana (BC)

A Combinação Bayesiana (Bayesian Combination - BC) (ROKACH, 2010a; BUNTINE, 1992) usa a probabilidade posterior do classificador em conjunto com a acurácia do mesmo α . Esta abordagem é descrita pela Equação 3.5:

$$\Psi_{c_t} = P(\alpha_{c_t} | S_{val}) \times P(w_i | x) \quad (3.5)$$

onde $P(\alpha_{c_t} | S_{val})$ é a acurácia do t ésimo classificador c_t estimado numa base de validação S_{val} , e $P(w_i | x)$ é a probabilidade posterior de c_t para a classe w_i , dada a instância x . Após o somatório de todas as saídas dos classificadores, o padrão a ser rotulado terá sua classe decidida de acordo com a pontuação associada para cada classe w_i , sendo que a classe com a maior pontuação é a decisão final do conjunto.

3.1.2.5 Valor de Poder (PV)

Em (LI; HOU; YIN, 2014), os autores propuseram o Valor de Poder (Power Value - PV) que é baseado em testes empíricos para definir o peso dos classificadores como uma variação de WMV. Esta abordagem é descrita na Equação 3.6:

$$\Psi_{c_t} = \frac{\alpha_{c_t}^{pv}}{\sum_{j=1}^T \alpha_{c_j}^{pv}} \quad (3.6)$$

onde pv é um valor escolhido empiricamente. Os autores avaliaram 13 valores diferentes para pv e concluíram que $pv = 10$ produzia o melhor resultado. Se $pv = 0$, então Ψ_{c_t} é definido apenas como a média da acurácia.

3.1.2.6 Métodos inspirados pela autoridade de um classificador (RSWV, BWWV e QBWWV)

$$\Psi_{c_t} = \log \left(\frac{a_{c_t}}{1 - a_{c_t}} \right) \quad (3.7)$$

As abordagens Voto Ponderado por Redimensionamento (Re-scaled weighted vote - RSWV), Melhor-Pior Voto Ponderado (Best-worst weighted vote - BWWV), e Melhor-Pior Voto Ponderado Quadrático (Quadratic best-worst weighted vote - QBWWV) estimam a primeiramente a autoridade de um classificador a_{c_T} . Essa autoridade é usada para estimar o peso final dos classificadores usando a Equação 3.7, a qual é inspirada em (KUNCHEVA, 2014). Alguns classificadores podem receber peso zero no caso de terem acurácia pobre (neste caso acontece na prática uma poda do conjunto original de classificadores e não a combinação de todos os membros do conjunto). Estes métodos estão apresentados como descrito em (MORENO-SECO et al., 2006).

Voto Ponderado por Redimensionamento (RSWV)

A primeira abordagem, RSWV, estima a autoridade dos classificadores como descreve a Equação 3.8:

$$a_{c_t} = \max \left\{ 0, 1 - \frac{M \times \epsilon_{c_t}}{V \times (M - 1)} \right\} \quad (3.8)$$

onde M é o tamanho do conjunto de classes W , V é o tamanho do conjunto de validação S_{val} , e ϵ_{c_t} é o número de erros do classificador c_t . Se $a_{c_t} \leq \frac{1}{M}$, então o peso do classificador é zero, caso contrário o peso do classificador é proporcional a acurácia.

Melhor-Pior Voto Ponderado (BWWV)

A abordagem BWWV assinala zero ao pior classificador e um ao melhor, após estimar a acurácia individual dos classificadores utilizando uma base de validação S_{val} . Como consequência, o pior classificador tem seu voto desconsiderado. Os classificadores restantes são ponderados linearmente dentro do limite apresentado. A Equação 3.9 apresenta esta abordagem:

$$a_{c_t} = 1 - \frac{\epsilon_{c_t} - \epsilon_{BEST}}{\epsilon_{WORST} - \epsilon_{BEST}} \quad (3.9)$$

onde $\epsilon_{WORST} = \max_T \{\epsilon_{c_t}\}$ e $\epsilon_{BEST} = \min_T \{\epsilon_{c_t}\}$.

Melhor-Pior Voto Ponderado Quadrático (QBWWV)

A abordagem é inspirada por BWWV, mas assinala ainda mais peso para os classificadores mais acurados. Esta abordagem é apresentada pela Equação 3.10:

$$a_{c_t} = \left(\frac{\epsilon_{WORST} - \epsilon_{c_t}}{\epsilon_{WORST} - \epsilon_{BEST}} \right)^2 \quad (3.10)$$

Esta seção apresentou métodos de combinação encontrados na literatura com destaque aos métodos ponderados estaticamente. Como é possível observar, quando trata-se de considerar apenas o classificador individualmente, o critério preferido pelos autores é a acurácia (LIU et al., 2017). Portanto, a diferença entre os classificadores quanto a como ele influencia a decisão final do conjunto é baseada apenas no seu desempenho individual de reconhecimento. Porém, medir a influência dos classificadores considerando a diferença deste para os demais não é uma tarefa simples, isto é, a diferença é um atributo compartilhado, pois usa outro classificador ou um conjunto de classificadores para ser estimada. Se um método é capaz de utilizar a diferença dos classificadores de forma complementar ao conjunto, as interações entre os classificadores devem ser consideradas em métodos de fusão.

3.2 Seleção Estática de Classificadores

A Seleção de Classificadores (Ensemble Selection) considera que dado um pool de classificadores, o uso de um subconjunto de classificadores pode ser mais acurado que o uso de todos os classificadores. A fase de seleção pode ser estática ou dinâmica (AZIZI; FARAH, 2012; Britto Jr.; SABOURIN; OLIVEIRA, 2014). A seleção estática

de classificadores envolve a escolha de classificadores na fase de treinamento, portanto, este conjunto não é atualizado durante a fase de teste. A seleção dinâmica, ao contrário, seleciona subconjuntos na fase de teste, possibilitando diferentes subconjuntos para cada padrão novo a ser rotulado. Geralmente, a seleção estática apresenta como vantagem uma boa taxa de reconhecimento a um custo menor de recursos, enquanto que a dinâmica utiliza mais recursos para fornecer uma taxa de reconhecimento melhor, pois ajusta-se a necessidade de cada padrão a ser rotulado.

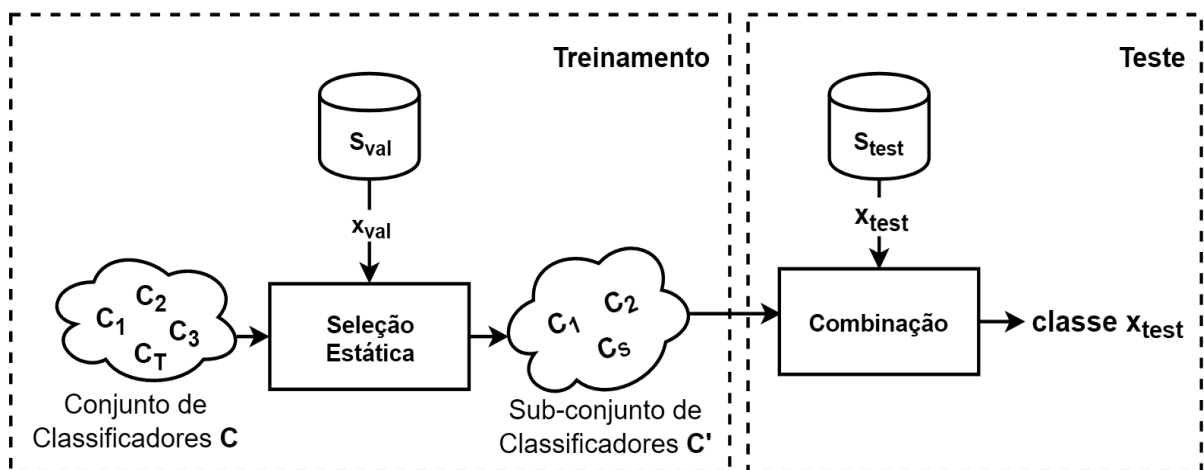


Figura 5 – Sistema de Múltiplos Classificadores com Seleção Estática de Classificadores.

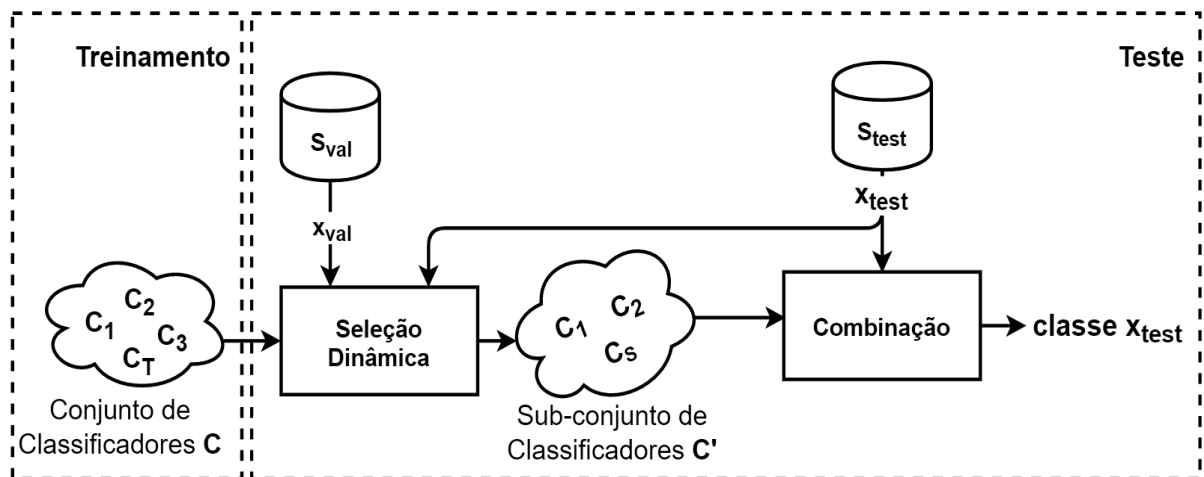


Figura 6 – Sistema de Múltiplos Classificadores com Seleção Dinâmica de classificadores.

Estes tipos de seleção estão representados na Figura 5 e na Figura 6. Na seleção estática (Figura 5) um conjunto inicial de classificadores passa por um processo de escolha do subconjunto mais apto para a classificação na fase de treinamento. Portanto, o conjunto de classificadores C passa por um procedimento de redução, onde $|C'| \leq |C|$ e $C' \subseteq C$. Este subconjunto de classificadores é fixo para a classificação de toda a base de teste. Na

seleção dinâmica (Figura 6) a escolha do subconjunto de classificadores mais aptos acontece na fase de teste. O objetivo desta mudança é considerar que sendo cada instância de teste diferente, há um subconjunto de classificadores mais apropriados para a classificação individual delas. A seleção dinâmica pode escolher um classificador para cada instância de teste ou pode escolher o melhor subconjunto para a instância de teste como pode ser observado em (CRUZ; SABOURIN; CAVALCANTI, 2018). Nesta proposta a seleção estática é abordada com ênfase em estratégias gulosas (*greedy*).

Dentre os trabalhos pioneiros na seleção estática está a Superprodução e Escolha (Overproduce and Select ou Overproduce and Choose), sendo descrito no trabalho de Partridge e Yates (PARTRIDGE; YATES, 1996). Na Figura 7 Nabiha et. al. (NABIHA; NADIR; MOKHTAR, 2011) propuseram um esquema para ilustrar o método de Superprodução e Escolha. Na fase de superprodução de conjunto, é comum o uso de Bagging, Boosting ou Random Subspaces, que são algoritmos clássicos de geração de classificadores. Alguns classificadores (*N Best*) ou apenas um classificador (*Single Best*) são definidos na escolha de classificadores, também chamada de fase de seleção. Para o caso de haver mais de um classificador na fase de escolha, há a necessidade de um método de combinação do voto desses classificadores. Para encerrar, o método como um todo passa por uma avaliação das estratégias adotadas na fase de avaliação de desempenho.

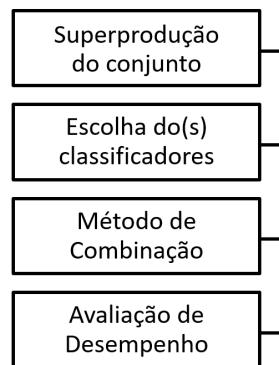


Figura 7 – Superprodução e Escolha (NABIHA; NADIR; MOKHTAR, 2011).

Considerando o modelo de superprodução e escolha (Figura 7), na fase de escolha dos classificadores tem-se pela heurística o classificador mais promissor, também conhecido como o Melhor Classificador (*Single Best*) ou a combinação de um conjunto de classificadores (*N best*). No caso da seleção de um subconjunto de mais de um classificador, *N best* refere-se a um número fixo ou não.

3.2.1 Estratégias Gulosas ou Gananciosas

Ainda sobre a seleção de um número flexível de classificadores existem as estratégias gulosas ou gananciosas. Estas estratégias tem por finalidade a adição ou remoção de classificadores a um subconjunto pela avaliação dos classificadores presentes no pool inicial. Dentre os aspectos principais a serem tratados por estratégias gulosas, destacam-se a i) quantidade de classificadores determinada de forma fixa ou flexível, ii) a medida que avalia qual é o classificador candidato a ser incluído ou removido do subconjunto final C' e iii) disponibilidade de dados para a avaliação dos classificadores. Este último é um dilema generalizado na área, então o uso de uma base de validação é recomendada para evitar o *overfitting* e também a criação de um sistema de múltiplos classificadores. A medida de avaliação dos classificadores pode ser uma função f para avaliar um classificador candidato c_i em união com o subconjunto C' .

Dentre as várias funções apresentadas na literatura, as que foram criadas especificamente para estratégias gananciosas de seleção de acordo com (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012) são o complemento, a concorrência e a minimização da distância da margem (*margin distance minimization*), as quais tendem a ser baseadas em acurácia, diversidade ou ambas. Quanto ao número de classificadores, cada estratégia possui o seu critério, porém, o mais simples é fixar os n classificadores que se destacam mediante o uso de uma função de avaliação, que por sua vez apresenta-se comumente como a estimativa de acurácia.

Complementariedade

A complementariedade de um classificador c_i com respeito ao subconjunto C' e um conjunto de validação S_{val} é calculado como apresentado pela Equação 3.11:

$$COM_{S_{val}}(c_k, C') = \sum_{i=1}^N (I(y_i = c_k(x_i) \ E \ y_i \neq C'(x_i))), \quad (3.11)$$

onde $I(\text{verdade})=1$, $I(\text{false})=0$ e $S_{val}(x_i)$ é a classificação da instância x_i pertencente ao conjunto de validação S_{val} . Desta forma, a complementariedade (COM) de um modelo em relação ao subconjunto é a quantidade de exemplos em S_{val} que o classificador candidato c_i classifica corretamente os exemplos em que o subconjunto C' não é capaz. Este complemento se assemelha a relação N_{10} vista na Seção 2.4, porém, N_{10} é o complemento de cada classificador c_i para outro classificador c_j , enquanto que $COM_{S_{val}}(c_i, C')$ procura avaliar o complemento de c_i em relação ao subconjunto C' .

Concorrência

A concorrência de um classificador c_i em relação ao subconjunto C' é dada pela Equação 3.12:

$$\begin{aligned} CON_{S_{val}}(c_k, C') = \sum_{i=1}^N & (-2 * I(y_i \neq c_k(x_i) \ E \ y_i \neq C'(x_i)) \\ & + 2 * I(y_i = c_k(x_i) \ E \ y_i \neq C'(x_i)) \\ & + I(y_i = c_k(x_i) \ E \ y_i = C'(x_i))) \end{aligned} \quad (3.12)$$

Minimização da margem de distância

A minimização da margem de distância é uma abordagem como o próprio nome diz, baseada em distância. Porém, alguns passos são apresentados antes de apresentar como funciona esta distância. Primeiro, para cada classificação correta de c_i em S_{val} , tem-se 1 ou -1 caso contrário. As respostas de cada classificador c_i são armazenadas em seu respectivo vetor $classif_i$. O vetor $Classif_{C'}$ armazena a média dos vetores individuais $classif_i$, logo $\frac{1}{|C'|} \sum_{i=1}^{|S|} classif_i$. Quando o subconjunto classifica corretamente todas as instâncias do conjunto de validação, o vetor correspondente encontra-se no primeiro quadrante no hiperplano N-dimensional. Então, o objetivo é reduzir a distância $d(o, Classif)$ onde d é a distância Euclidiana e o é um vetor pré-definido que encontra-se no primeiro quadrante. A margem $MMD_{S_{val}}(c_k, C')$ do classificador c_k em relação ao subconjunto C' e um conjunto de validação S_{val} é representada pela Equação 3.13:

$$MMD_{S_{val}}(c_k, C') = d(o, \frac{1}{|C'| + 1} (classif_k, Classif_{C'})) \quad (3.13)$$

Seleção Focada em Diversidade

A medida Seleção Focada em Diversidade (*Focused Selection Diversity - FSD*) considera 4 diferentes situações semelhantes a N_{10} , N_{01} , N_{11} e N_{00} apresentadas como base para estimar as medidas de diversidade na Seção 2.4. Como essa medida avalia o voto do classificador candidato c_i em relação ao subconjunto C' tem-se:

1. N_{10} : c_i está correto enquanto que C' não está;
2. N_{01} : c_i está incorreto enquanto que C' está correto;
3. N_{11} : c_i está correto e C' também está;

4. $N_{00:c_i}$ está incorreto e C' também está;

Essas possibilidades foram combinadas na Equação 3.14:

$$\begin{aligned}
 FSD_{S_{val}}(c_k, C') = \frac{1}{|C'|} \sum_{i=1}^N & (NT_i * I(y_i = c_k(x_i) \ E \ y_i \neq C'(x_i)) \\
 & - NF_i * I(y_i \neq c_k(x_i) \ E \ y_i = C'(x_i)) \\
 & + NF_i * I(y_i = c_k(x_i) \ E \ y_i = C'(x_i)) \\
 & - NT_i * I(y_i \neq c_k(x_i) \ E \ y_i \neq C'(x_i))),
 \end{aligned} \tag{3.14}$$

onde NT é o número de classificadores em C' que classificam corretamente a instância i do conjunto S_{val} e NF é o contrário.

Em (RUTA; GABRYS, 2005), foram apresentadas as estratégias mais populares de seleção gananciosa, a *Forward Search* e a *Backward Search*. A seguir, são apresentadas estas e outras estratégias que se destacam na literatura.

3.2.1.1 Forward Search (FS)

A estratégia *Forward Search* (FS) adiciona inicialmente a um subgrupo vazio ($|C'| = \emptyset$) o *Single Best*. A partir daí começa a avaliar o conjunto de classificadores aos pares, visando a redução do erro de C' . A cada iteração é escolhido o par de classificadores que mais contribui a C' . O algoritmo finaliza quando não há mais contribuições a serem feitas pelos pares avaliados. Alternativamente, a adição dos classificadores pode ser feita individualmente como apresentado no Algoritmo 3 semelhante ao descrito em (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012).

3.2.1.2 Backward Search (BS)

A estratégia *Backward Search*, ao contrário de *Forward Search* busca o melhor subconjunto C' iniciando-o com o conjunto original C e diminuindo a cada iteração. O critério de exclusão geralmente é a acurácia do subconjunto C' quando removido o par que maximize a acurácia do subconjunto C' . O critério de parada neste caso é determinado pela indiferença na remoção de qualquer par de classificadores contido em C' . Tanto *Backward Search* quanto *Forward Search* tem complexidade $O(T^3)$, onde T é o tamanho do pool inicial. Assim como para *Forward Search*, a adição dos classificadores pode ser feita individualmente como apresentado em (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012) e pelo Algoritmo 4.

Algoritmo 3: FS (C, f_{FS}, S_{val})

Entrada: Conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$ **Entrada:** Função de avaliação (f_{FS})**Entrada:** Base de validação (S_{val})

```

1  $C' \leftarrow \emptyset$ ;
2 repetir
3    $A_{C'} \leftarrow$  acurácia do subconjunto  $C'$ ;
4   para cada  $c_i \in C$  faça
5      $A_i \leftarrow$  acurácia ( $f_{FS}$ ) de  $c_i \cup C'$  ;
6   fim
7    $A_M \leftarrow$  o valor da maior acurácia;
8    $c_M \leftarrow$  o classificador de maior acurácia;
9   se  $A_M \geq A_{C'}$  então
10     $C' \leftarrow c_M \cup C'$ ;
11     $C \leftarrow C \setminus c_M$ ;
12  fim
13 até que  $A_M < A_{C'}$ ;
14 retorne  $C'$ ;

```

Algoritmo 4: BS (C, f_{BS}, S_{val})

Entrada: Conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$ **Entrada:** Função de avaliação (f_{BS})**Entrada:** Base de validação (S_{val})

```

1  $C' \leftarrow C$ ;
2 repetir
3    $A_{C'} \leftarrow$  acurácia do subconjunto  $C'$ ;
4   para cada  $c_i \in C'$  faça
5      $A_i \leftarrow$  acurácia ( $f_{BS}$ )  $C' \setminus c_i$ ;
6   fim
7    $A_M \leftarrow$  o valor da maior acurácia;
8    $c_M \leftarrow$  o classificador de maior acurácia;
9   se  $A_M > A_{C'}$  então
10     $C' \leftarrow C' \setminus c_i$ ;
11  fim
12 até que  $A_M < A_{C'}$ ;
13 retorne  $C'$ ;

```

3.2.1.3 Kappa Pruning

Na abordagem *Kappa Pruning* (MARGINEANTU; DIETTERICH, 1997), um subconjunto C' inicia-se vazio e é acrescentado de um classificador c_i a cada iteração. O classificador c_i é adicionado em C' até que C' atinja um tamanho fixo de N classificadores. Originalmente os autores escolheram os classificadores mais promissores fixando empiricamente N em tamanhos diferentes (80%, 60%, 40% e 20%) e comparando os resultados obtidos separadamente para cada tamanho para obter uma conclusão geral. Eles concluíram que a estratégia era capaz de reduzir o conjunto sem perda na taxa de reconhecimento para a maior parte dos problemas, sugerindo que a redução entre 60% a 80% apresentava os resultados mais atrativos. Este método é até hoje comparado a alternativas mais complexas de seleção, como o Algoritmo Genético proposto em (CAVALCANTI et al., 2016) que escolhe um número N de classificadores de modo flexível inspirados pela combinação de diferentes medidas de diversidade.

3.2.1.4 Seleção de Subconjunto de Classificadores Regulada por Diversidade (DREP)

Uma estratégia de seleção gananciosa recente é a Seleção de Subconjunto de Classificadores Regulada por Diversidade (*Diversity regularized ensemble pruning* - DREP) (LI; YU; ZHOU, 2012). Esta estratégia tem inicialmente um subconjunto C' vazio. Em seguida acrescenta o *Single Best* - obtido a partir da análise da maior acurácia individual observada num conjunto de validação S_{val} . As próximas iterações são feitas somente nos classificadores que não pertencem a C' . Um a um, esses classificadores são avaliados quanto a diversidade dele para C' , inspirado pelas medidas de avaliação apresentadas nesta seção. Em seguida, uma porção deles é avaliada de forma a obter o classificador que mais complementa as decisões de C' . Caso esse classificador candidato complemente o subconjunto ele é acrescentado a ele. Caso contrário, o algoritmo pára. Esta estratégia é bastante competitiva com métodos mais complexos, como o Algoritmo Genético proposto em (CAVALCANTI et al., 2016). A grande vantagem que esta estratégia gulosa apresenta em relação aos clássicos *Forward Search* e *Backward Search* é implementar a troca (*trade-off*) entre a diversidade e a acurácia no seu processo de decisão.

Quanto aos problemas que as estratégias gulosas apresentam, dois se sobressaem: a disponibilidade dos dados para a avaliação dos subconjuntos e a função que define a permanência ou exclusão do subconjunto final. O primeiro problema tem como soluções populares o uso da mesma base de treinamento para avaliação, ou o uso de uma base de validação para evitar o *overfitting*. Quanto ao segundo problema, a função de avaliação, a literatura

Algoritmo 5: DREP (C, S_{val}, p)

Entrada: Conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$
Entrada: Base de validação (S_{val})
Entrada: Proporção de candidatos mais diversificados a ser avaliados (p)

- 1 $C' \leftarrow \emptyset$;
- 2 $c_i \leftarrow$ o classificador com menor erro em S_{val} ;
- 3 $C' \leftarrow c_i$;
- 4 $C \leftarrow C \setminus c_i$;
- 5 **repetir**
- 6 **para** cada $c_i \in C$ **faça**
- 7 $d_i \leftarrow$ a discordância entre c_i e C' ;
- 8 **fim**
- 9 ordenar os classificadores c_i em C pela maior discordância d_i ;
- 10 $\Gamma \leftarrow$ os primeiros $\lceil p \cdot |C| \rceil$ classificadores da lista ordenada;
- 11 $c_i \leftarrow$ o classificador em Γ que mais reduz o erro do subconjunto C' em S_{val} ;
- 12 $C' \leftarrow c_i \cup C'$;
- 13 $C \leftarrow C \setminus c_i$;
- 14 **até que** o erro de C' em S_{val} não possa ser reduzido;
- 15 **retorne** C' ;

apresenta escolhas baseadas em duas categorias principais (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012): desempenho (acurácia) ou diversidade. Tanto *Forward Search* quanto *Backward Search* geralmente consideram somente a acurácia de um classificador candidato $c_i \in C$ quando adicionado ao subconjunto C' , logo $c_i \cup C'$ é avaliado, adicionando o classificador candidato c_i que maximize a acurácia de C' . Entretanto, a abordagem DREP incentiva a escolha dentre os mais divergentes candidatos classificadores $c_i \in C''$ aquele que mais contribui para a acurácia de C' . Para o cálculo da diversidade, a literatura apresenta as medidas pareadas e não pareadas vistas na Seção 2.4. DREP opta por calcular somente a diversidade entre o candidato c_i em relação a C' .

3.2.1.5 Aggregation Ordering in Bagging (AGOB)

Esta abordagem é proposta e descrita em (MARTÍNEZ-MUÑOZ et al., 2004) como uma abordagem específica para conjunto de classificadores criados por *Bagging*, entretanto, a solução pode ser usada para outros métodos de criação. O autor propôs 3 variações para este algoritmo. A primeira, começa com a adição do classificador de maior desempenho a um subconjunto e depois cada classificador que adicionado ao subconjunto proporciona a maior acurácia para o subconjunto é adicionado a lista, semelhante ao que o método *Forward Search* faz, porém, mais sofisticado. A segunda proposta apenas muda o critério de adição dos classificadores, passando ao que mais complementa as instâncias as quais

o subconjunto ($COM_{S_{val}}(c_k, C')$) tem dificuldade de acertar. A terceira proposta muda o critério de adição para a minimização da margem de distância ($MMD_{S_{val}}(c_k, C')$). Os autores concluem que a complementariedade é uma medida mais atrativa para selecionar um número arbitrário de classificadores pela estratégia por eles proposta.

3.2.1.6 Pruning in Ordered Bagging ensembles (POBE)

Em (MARTÍNEZ-MUÑOZ; SUÁREZ, 2006) os autores propuseram usar a função de *minimização da distância da margem* para ordenar os classificadores de um ensemble. Depois de ordenar, apenas os classificadores mais promissores são escolhidos, sendo que o tamanho do conjunto é definido empiricamente. Os testes conduzidos pelos autores mostraram que conjuntos com 15 a 30% do tamanho total do conjunto produziam um resultado quase ótimo quando avaliado numa base de treinamento. Entretanto, esta observação não se repetiu na avaliação da abordagem numa base de teste.

3.3 Considerações finais

Este capítulo apresentou alguns métodos da literatura que estão associadas a técnicas de seleção gananciosa, também conhecidas como seleção gulosa (*Greedy Selection*). Foram apresentados os algoritmos *Forward Search*, *Backward Search*, *Kappa Pruning*, AGOB, POBE e o recente DREP. Porém, outras abordagens podem ser encontradas em (PARTALAS; TSOUMAKAS; VLAHAVAS, 2012; MAO et al., 2011). Um aspecto interessante considerado nas abordagens mencionadas é a complementariedade de um classificador (em relação a um subconjunto previamente selecionado) como um critério para a seleção.

É importante mencionar que AGOB e POBE foram propostos com a intenção de ponderar classificadores gerados pela técnica de Bagging como uma alternativa ao Boosting. Entretanto, se o Boosting é encarado como apenas um método de geração de classificadores que ao final do processo não pondera o voto dos classificadores, tem-se um método de geração induzido pela dificuldade de classificar algumas instâncias que pode ser usado em métodos de seleção gananciosa com a vantagem de ser um método de geração que induz os classificadores a serem complementares as dificuldades encontradas pelo ensemble.

As técnicas de fusão apresentadas têm em comum o uso da acurácia individual dos classificadores como parte do processo de ponderação do voto dos classificadores. A

importância do classificador recebe a acurácia do classificador como entrada e ajusta o peso do voto do classificador por alguma regra ou fórmula. Estes métodos apresentados na literatura não consideram a diversidade entre os classificadores, critério normalmente usado como parte do processo de seleção de um subconjunto de classificadores.

Ambas as técnicas apresentadas neste capítulo constituem concorrentes das abordagens propostas no próximo capítulo. A ponderação do classificador difere dos métodos de fusão apresentados nesta seção no quesito da análise do relacionamento de diversidade entre os classificadores. Além do método de fusão, uma estratégia para a seleção de classificadores é proposta, usando a importância dos classificadores como um dos critérios para a composição de um subconjunto.

4 Metodologia Proposta

Neste capítulo está descrita a metodologia proposta de combinação estática de classificadores e de seleção estática de classificadores. As abordagens de combinação e de seleção compartilham características comuns, portanto, esse capítulo é iniciado com a apresentação de um SMC como uma rede de classificadores. Em seguida, descreve-se a fusão ponderada estática. Por fim, é apresentada a abordagem de seleção, a qual foi desenvolvida inspirada pelo método de fusão.

4.1 Rede de Classificadores

A rede de classificadores recebe um destaque nesta proposta por ser o núcleo do trabalho. Afinal, é a partir da análise desta rede que se discute a aplicabilidade da estimativa de importância de seus membros para a construção de métodos de seleção e de fusão.

Uma rede $G(V, E)$ é composta - como apresentada na seção de revisão de literatura - por um grupo de objetos representados por V e por um grupo E , o qual representa as relações pareadas entre estes objetos. O conjunto de objetos $V = C = \{c_1, c_2, \dots, c_T\}$ são portanto, aqui considerados como os classificadores obtidos por algum método de geração de pool enquanto que o conjunto E representa as relações pareadas, como medidas de diversidade, por exemplo. Desta forma, cada objeto $e(i, j) \in E$ representa uma pontuação referente às diferenças entre os classificadores c_i e c_j . Note que para medidas de diversidade entre classificadores que são simétricas, $e(i, j) = e(j, i)$, mas para relações assimétricas $e(i, j) \neq e(j, i)$.

Na Figura 8 está representada a criação e posterior análise de uma rede de classificadores. Para a representação da rede, o conjunto de classificadores C e o conjunto ϖ (representando relações pareadas) são necessários. A partir da rede de classificadores representada como um grafo G , passa-se a analisar a importância de cada um deles utilizando alguma medida de centralidade χ . O objetivo de analisar esta rede é obter um conjunto de pesos Ψ , que estima a pontuação de importância de cada membro C_i deste conjunto de classificadores.

A Rede de Classificadores é a entrada para um método de fusão e também para um método de seleção, os quais estão descritos nas seções seguintes.

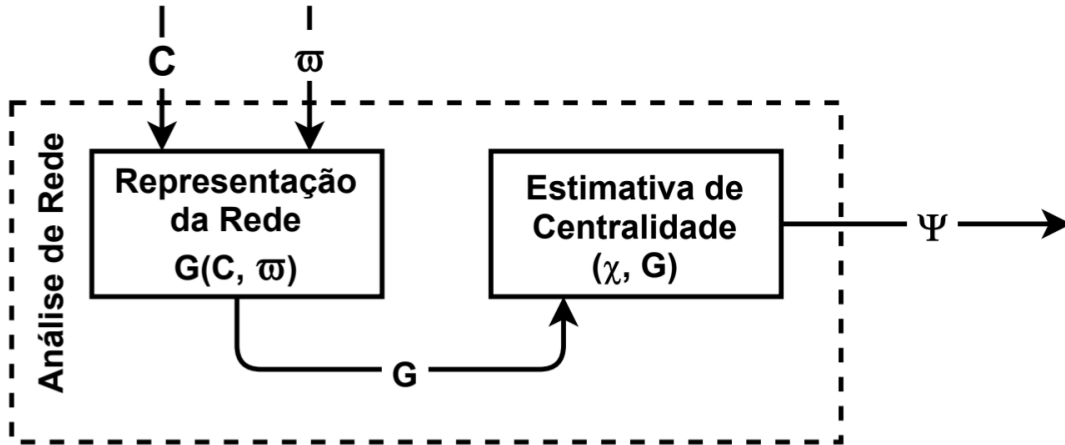


Figura 8 – Construção e Análise da Rede de Classificadores.

4.2 Fusão Estática dos Classificadores (CBF)

O método descrito para a fase de fusão de classificadores é nomeado Fusão Baseada em Centralidade (Centrality Based Fusion - CBF). A Figura 9 apresenta uma visão geral deste método de fusão ponderada e estática. Este método foi proposto em 3 etapas. Na primeira, um conjunto de classificadores C é gerado na fase de Produção, usando uma base de treinamento $S_{train} \in S$, na qual as instâncias x_{train} são selecionadas para compor T bases de treinamento de acordo com algum método de geração de pool da literatura. Na segunda fase, a diversidade entre os classificadores é estimada e armazenada em ϖ . Para isto, uma medida pareada, e.g., medida de diversidade precisa ser escolhida. A primeira e a segunda fase fornecem as entradas para o método de fusão proposto. Nesta terceira fase, a Rede de Classificadores é construída e representada como um grafo ponderado $G(V, E)$, onde o conjunto de vértices V representa o conjunto de classificadores C e o conjunto de arestas E representa a diversidade entre os classificadores, denotada por ϖ . Este grafo passa a ser analisado, com o objetivo de estimar a importância dos classificadores. Para isso, alguma medida de centralidade χ é empregada. Após a estimativa da importância dos membros de C , o peso sugerido pela centralidade determina a influência do classificador no processo de rotulação de objetos.

De maneira geral, CBF faz uso de um conjunto de dados S , o qual é dividido em três partes distintas, utilizadas para treino, validação e teste. A base de treinamento S_{train} provê instâncias x_{train} para a geração de um pool de classificadores C preferencialmente acurados e distintos. O conjunto C é utilizado para estimar a diferença entre os classificadores de maneira pareada adotando para tal uma base de validação S_{val} . O pool de classificadores C e as relações pareadas entre eles (ϖ) servem de entrada para a construção da rede de

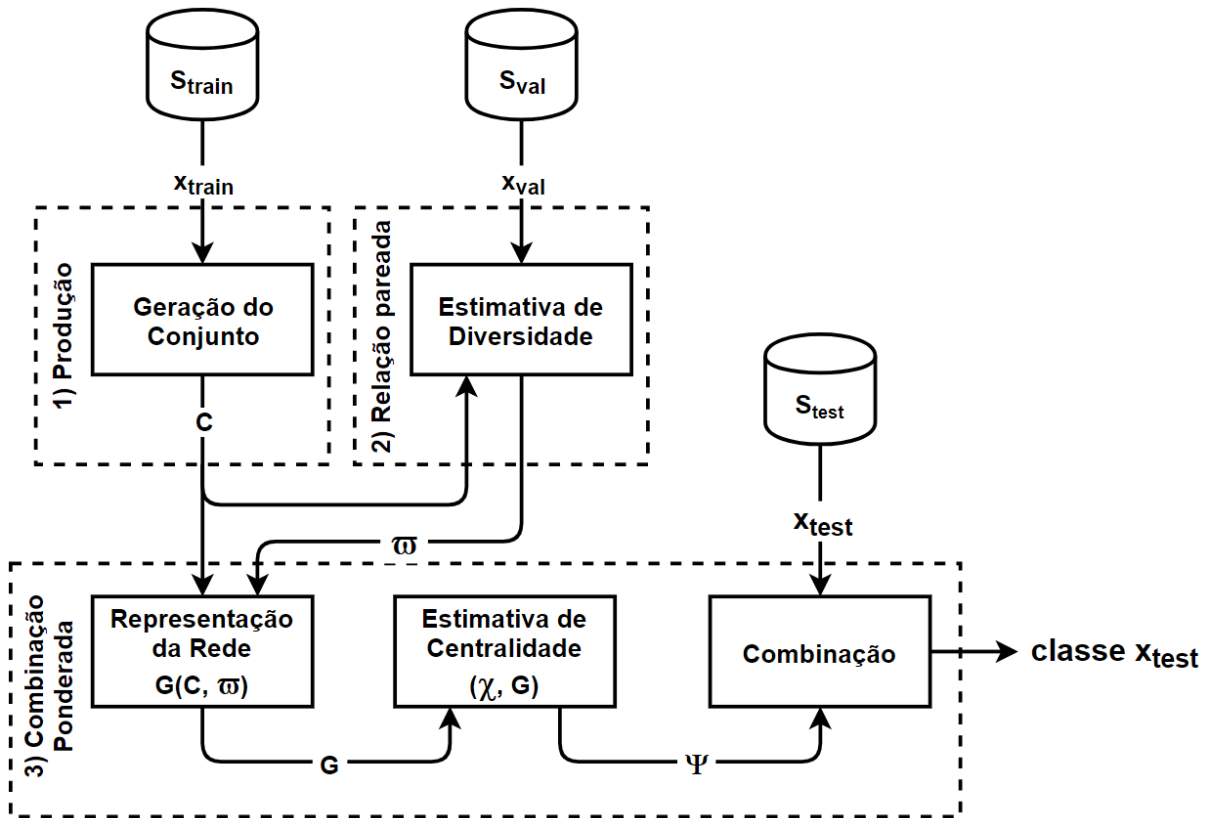


Figura 9 – Visão geral do método CBF.

classificadores, representada pelo grafo $G(C, \varpi)$, no qual vértices representam os classificadores, e as relações pareadas são representadas por arestas, logo $G(V, E) = G(C, \varpi)$. Note que, por existirem relações pareadas simétricas ou assimétricas, deve-se escolher cuidadosamente a medida de centralidade χ . Algumas fórmulas para relacionamentos simétricos e assimétricos foram apresentados na Seção 2.1.3. A medida de centralidade χ é usada para estimar a importância dos classificadores, portanto, depende do que as relações pareadas representam e como a centralidade interpreta estes relacionamentos. A pontuação referente a análise de importância do classificador é definida por Ψ . Finalmente, uma regra de combinação de classificadores é utilizada, atuando sobre a interpretação do peso dos classificadores. Desta forma, a instância de teste x_{test} é classificada. Cada fase é detalhada a seguir.

4.2.1 Produção

A fase de Produção gera um conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$, no qual cada membro representa uma função independente $c_t : R^n \rightarrow W$ que assinala um rótulo de classe $w_i \in W$ para uma instância $x \in R^n$, onde $W = \{w_1, w_2, \dots, w_M\}$. Os métodos mais populares de geração de classificadores estão descritos na Seção 2.3. Entretanto, a

abordagem nova apresentada, CBF, não está limitada a nenhum destes métodos.

4.2.2 Relação Pareada

A fase de estimativa de Relação Pareada serve para avaliar como os classificadores trabalham juntos. No caso de medidas de diversidade o objetivo é descobrir a diferença entre o par de classificadores. Assim, qualquer medida apresentada na Seção 2.4 como a Estatística Q, o Coeficiente de Correlação, o Desacordo, a Falta Dupla, ou a Estatística Kappa. Para estas medidas, o relacionamento é simétrico. Há também relacionamentos pareados como N_{10} , o qual expressa uma dependência. Esse tipo de relacionamento assimétrico sugere que a seta indique a dependência, portanto, a seta aponta para o classificador que está correto. Nesta fase, o conjunto de classificadores $C = \{C_1, \dots, C_t\}$ é usado para estimar um desses tipos de relação pareada (simétrica ou assimétrica) $\varpi(C_i, C_j)$ entre qualquer par de classificadores. Para isso, é indicado o uso de uma base de validação para evitar a sobreposição dos dados de treinamento com os de avaliação (*overfitting*).

4.2.3 Combinação Ponderada

A proposta de CBF difere de muitas abordagens encontradas na literatura em um aspecto principal: o relacionamento entre os classificadores é que determina a influência do classificador para o conjunto e, por consequência, é utilizado como peso em um sistema de fusão de classificadores. Nesta fase, o núcleo da proposta de fusão, é descrito por três passos distintos:

- *Representação da Rede de Classificadores*: formalmente, a rede de classificadores é formada por um conjunto de classificadores (C) e a relação pareada entre eles (ϖ). Um vértice representa um classificador do conjunto C enquanto que ϖ representa arestas direcionadas ou não-direcionadas. Para relações simétricas, tem-se arestas não-direcionadas, pois $\varpi(C_i, C_j) = \varpi(C_j, C_i)$. Já para as relações assimétricas, considera-se $\varpi(C_i, C_j) \neq \varpi(C_j, C_i)$ e, portanto, as arestas da rede devem ser direcionadas. Qualquer medida de diversidade apresentada na Seção 2.4 representa relações simétricas enquanto apenas as relações N_{01} e N_{10} são assimétricas. No caso de relações simétricas, o número possível de relações a partir do conjunto de classificadores é estimado pela combinação $C_{T,2}$ onde T é o tamanho do conjunto C e 2 refere-se ao relacionamento entre os classificadores, em par. Por exemplo, um comitê de $T = 100$ classificadores tem $C_{100,2} = 4950$ relações pareadas e simétricas. Por outro

lado, o número de relações assimétricas é estimado pela permutação (arranjo) $P_{T,2}$. Então para o mesmo tamanho de conjunto tem-se $P_{100,2} = 9900$ relações pareadas e assimétricas. Assim, encerra-se este passo com a saída de uma rede $G(V, E)$ que pode ser direcionada ou não-direcionada, na qual $V = C$ e $E = \varpi$.

- *Estimativa de Centralidade*: um grafo é uma entrada requerida para estimar a importância dos elementos na rede por meio de uma medida de centralidade. Medidas de centralidade clássicas como *Intermediação*, *Proximidade*, *Grau*, e *Autovetor* são apropriadas para analisar relações simétricas ou assimétricas, entretanto, ao considerar as equações apresentadas na Seção 2.1.3, estas são usadas neste trabalho para relacionamentos simétricos. Para os relacionamentos assimétricos, as medidas de centralidade *Grau de Entrada* e *Grau de Saída* são apropriadas. Estas exigem mais cuidado, pois a escolha delas depende da direção do relacionamento que pretende-se analisar. O parâmetro χ refere-se a qualquer medida de centralidade, como as clássicas, descritas na Seção 2.1.3. Qualquer destas medidas sugerem a importância do classificador por uma pontuação proveniente da análise da posição do classificador na rede. Para algumas medidas, a relação com mais peso pode ser mais importante, mas para outras, o menor peso nos relacionamentos é que sugerem maior importância. Deste modo, é interessante a adaptação do peso dos relacionamentos em função da medida de centralidade que será utilizada.
- *Combinação*: neste estágio, o método de fusão CBF já tem estimada a pontuação referente a importância dos classificadores (Ψ). Passa-se então para o passo da classificação respeitando a influência de cada classificador C_i determinado pelo peso Ψ_i . Assim, uma instância de teste x_{test} recebe o voto de todos os classificadores. A classe que recebe a maior pontuação é a classe a qual o conjunto acredita ser a correta. Portanto, os votos atribuídos a cada classe são descritos como simplesmente um esquema de ponderação como WMV (Equação 3.2) e a decisão pela classe correta segue como descrito pela Equação 3.1.

4.2.4 O Algoritmo CBF

Na Figura 9 foi apresentada uma visão geral do método proposto de fusão. Alguns aspectos podem ter sido simplificados para facilitar o entendimento. Portanto, uma descrição detalhada é importante. Nesta seção é apresentado o algoritmo 6 referente ao método de fusão proposto.

Algoritmo 6: CBF(C, ϖ, χ)

Entrada: Conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$
Entrada: Relação pareada (ϖ), onde $\varpi(c_i, c_j)$
Entrada: Medida de centralidade (χ)
Entrada: Normalização (τ)
Entrada: Função de inversão (φ)

- 1 $V \leftarrow C$;
- 2 $E \leftarrow \varpi$;
- 3 **se** $\varphi = VERDADEIRO$ **então**
- 4 **para** $E_i \in E$ **faça**
- 5 $E_i = 1/E_i$;
- 6 **fim**
- 7 **se** $\tau = VERDADEIRO$ **então**
- 8 $E = normalizar(E)$;
- 9 **fim**
- 10 **fim**
- 11 Construir a rede $G(V, E)$;
- 12 $\Psi \leftarrow$ Estimar centralidade (χ, G);
- 13 **retorne** Ψ ;

Como é possível notar, primeiramente, um conjunto de classificadores C é requerido como entrada. O parâmetro ϖ é responsável pelas relações pareadas entre os classificadores, ponderada de acordo com a divergência entre eles. Para este propósito, as relações pareadas apresentadas na Seção 2.4 podem ser utilizadas. No caso de relações simétricas, o número de relações é definido como uma combinação $C_{T,2}$ onde T refere-se ao tamanho do pool C e 2 é devido a combinação par a par.

As medidas de centralidade χ podem ser as seguintes para as relações simétricas (utilizando as fórmulas apresentadas na Seção 2.1.3): *Intermediação*, *Proximidade*, *Grau*, e *Autovetor*, enquanto *Grau de Entrada* ou *Grau de Saída* são indicadas para a análise de relacionamentos assimétricos. A normalização τ e a função de inversão φ são usadas para adaptar a rede para a análise correta a partir das medidas de diversidade. Após preencher estes requerimentos, o pool de classificadores C é utilizado para formar o conjunto de vértices V na linha 1. Uma medida pareada ϖ como as medidas de diversidade, definem as arestas E entre pares de classificadores (linha 2). Nas linhas 3 - 7 a função para inverter o valor de normalização da rede devem ser aplicados para que algumas medidas de centralidade possam estimar corretamente a importância do classificador, pois elas baseiam-se na quantidade e no peso das arestas que unem dois classificadores. Isto é explicado mais tarde em detalhes ainda nesta seção. O processo de normalização (linhas 7 - 9) tem por função evitar valores negativos ou nulos em E , principalmente se a função de

inverter o peso das arestas foi utilizado (neste caso é sugerido o processo de normalização). O grafo $G(V, E)$ passa a ser construído (linha 11), sendo referenciado neste documento como a rede de classificadores. Para obter o peso do classificador Ψ_i , que representa a pontuação associada a importância do classificador na rede, uma medida de centralidade χ escolhida realiza esta estimativa, para isso necessita do grafo G (linha 12). A pontuação de importância do classificador Ψ provê o peso do classificador utilizado no processo de fusão de classificadores. Portanto, este método difere das abordagens comumente encontradas na literatura ao analisar as relações entre os classificadores como um critério de ponderação dos mesmos no lugar de simplesmente operar na acurácia individual dos membros do grupo.

A Tabela 2 apresenta a relação entre o peso da aresta e como ele influencia a estimativa da centralidade. Por exemplo, a *Centralidade de Grau* atribui uma pontuação de centralidade mais alta se o somatório das arestas ligadas a um vértice é maior que o somatório das arestas que estão ligadas a cada um dos demais. Portanto, a *Centralidade de Grau* considera mais central os vértices cujas arestas tem maior peso. Já para a *Centralidade de Intermediação*, quanto menor o peso das arestas ligadas a um classificador, maior será a chance deste participar dos geodésicos da rede e por consequência estar entre os vértices mais centrais na rede.

Para facilitar o entendimento da relação entre a medida de centralidade e a diversidade representada na rede, é apresentada a Tabela 3. Por exemplo, para a medida de diversidade *Falta Dupla*, que assinala uma pontuação baixa para pares de classificadores que evitam estar incorretos para os mesmos exemplos avaliados, tem-se a seguinte situação: i) As centralidades de *Grau* e *Autovetor* consideram que quanto maior o peso da aresta mais importante é o relacionamento, portanto, sendo a *Falta Dupla* pior quanto maior é o peso, logo ela deve ser adaptada. Esta adaptação consiste em inverter o peso dos relacionamentos e depois normalizar. Desta forma, as centralidades mencionadas atribuirão mais peso a classificadores cujos relacionamentos evitam a *Falta Dupla*; ii) As centralidades de *Intermediação* e *Proximidade* - que são baseadas em geodésico - consideram arestas de peso baixo mais interessantes que as de peso alto. Neste caso, está em conformidade com *Falta Dupla*. Diante deste cenário, caso a representação de rede seja feita por *Falta Dupla* e a estimativa de importância feita por *Centralidade de Grau* ou *Autovetor*, o parâmetro φ deve ser *VERDADEIRO*. Este parâmetro é muito importante para auxiliar no processo de estimativa de importância, portanto, a relação pareada representada na Rede de Classificadores e a medida de centralidade devem ser consideradas juntas no

ajuste dos parâmetros de CBF.

Tabela 2 – Interpretação da medida de centralidade quanto ao peso da aresta. (\downarrow) Quanto menor o valor do peso da aresta, maior é o valor da métrica. (\uparrow) Quanto maior o valor do peso da aresta, maior é o valor da métrica.

Medida de Centralidade	Intermediação	Proximidade	Grau	Autovetor
Valor da métrica é alto	\downarrow	\downarrow	\uparrow	\uparrow

Tabela 3 – Diversidade expressa como um valor alto (\uparrow) ou baixo (\downarrow) dependendo da medida de centralidade que avalia a importância dos membros da rede.

Centralidade	DF	CC	Dis	QS	KS
Grau	\uparrow	\uparrow	\downarrow	\uparrow	\uparrow
Intermediação	\downarrow	\downarrow	\uparrow	\downarrow	\downarrow
Proximidade	\downarrow	\downarrow	\uparrow	\downarrow	\downarrow
Autovetor	\uparrow	\uparrow	\downarrow	\uparrow	\uparrow

Na Tabela 4 é apresentada a parametrização sugerida para a melhor utilização da abordagem CBF. Tomando o exemplo de uma rede de classificadores contendo *Falta Dupla* como representação dos relacionamentos, temos este cenário para a estimativa da Centralidade de Grau: o parâmetro φ recebe V devido DF ser alto quando dois classificadores estão frequentemente errados para os mesmos exemplos. Se a estimativa da centralidade de grau favorece classificadores cujo peso é maior (maior erro do par), o inverso do peso de DF é o peso correto para pontuar a importância dos classificadores no sentido de evitar erros. Sempre que $\varphi = V$ recomenda-se $\tau = V$. Para medidas de diversidade cujos valores podem ser nulos ou negativos, também é necessária a normalização. Comumente em uma matriz de adjacência, a ausência de relacionamento é representada por zero. Neste caso, se a medida de diversidade pode ter peso zero, deve-se tomar cautela e em caso de dúvida normalizar.

Para redes de relacionamento assimétrico, a relação entre os classificadores C_i e C_j é diferente de C_j e C_i . Assim, uma rede direcionada deve ser empregada. A medida de centralidade χ para redes direcionadas podem ser a centralidade de *Grau de Entrada* ou *Grau de Saída*, bem como as medidas de *Centralidade de Intermediação* e de *Proximidade*. No caso da relação N_{10} a direção da aresta aponta para o classificador correto, logo, a centralidade de *Grau de Entrada* é uma medida apropriada e simples, pois esta avalia a centralidade do classificador pelo somatório do peso das arestas que apontam para o classificador correto.

Tabela 4 – Parâmetros indicados para o uso de CBF como método de fusão. (V) significa VERDADEIRO enquanto (F) significa FALSO.

ϖ	χ	φ	τ
DF	Grau	V	V
	Intermediação	F	F
	Proximidade	F	F
	Autovetor	V	V
CC	Grau	V	V
	Intermediação	F	V
	Proximidade	F	V
	Autovetor	V	V
Dis	Grau	F	F
	Intermediação	V	V
	Proximidade	V	V
	Autovetor	F	F
QS	Grau	V	V
	Intermediação	F	V
	Proximidade	F	V
	Autovetor	V	V
KS	Grau	V	V
	Intermediação	F	V
	Proximidade	F	V
	Autovetor	V	V

4.3 Método de seleção estática (CBS)

As observações quanto à utilidade da rede para o método de fusão descrito na seção anterior levaram a crer que a rede de classificadores analisada por medidas de centralidade poderia contribuir para a proposta de uma estratégia de seleção. A estratégia descrita nesta seção tem como inspiração toda a teoria de redes apresentada para a confecção do método de fusão *CBF* e também é inspirada por estratégias gulosas de seleção estática de classificadores, as quais foram apresentadas na Seção 3.2.

Um dos problemas encontrados nas estratégias gulosas é a definição da função de avaliação do conjunto como critério para a inclusão de um classificador candidato a um subconjunto. Foram apresentadas algumas medidas para tal na Seção 3.2.1. Estas medidas encontradas na literatura, normalmente comparam somente o classificador candidato c_i com o subconjunto C' . No caso de uma instância ser classificada corretamente por quase metade do subconjunto de classificadores, o voto do classificador candidato pode afetar diretamente o voto do sistema e, por consequência, as medidas que usam esse padrão.

Considerando que o subconjunto de classificadores irá votar a classe de uma instância colhendo individualmente a opinião de todos, é apresentado um algoritmo que

considera a diversidade estimada de forma coletiva a partir da centralidade do classificador estimada em relações de diversidade/complementariedade do ensemble representado como uma rede. Uma segunda etapa para a escolha do classificador que irá fazer parte de C' é a coleta de votos dos classificadores com maior pontuação de centralidade. Estes, terão sua acurácia testada quando em união com o subconjunto C' e aquele que mais aumenta a acurácia do subconjunto é escolhido para fazer parte de C' em definitivo. Desta forma, o algoritmo 7 apresenta a Seleção de Classificadores baseada em Centralidade (*Centrality Based Selection - CBS*).

4.3.1 O Algoritmo CBS

O Algoritmo 7 apresenta uma estratégia de seleção gulosa que usa a análise da Rede de Classificadores para a compor um subconjunto de classificadores acurado. O algoritmo recebe como entrada um conjunto de classificadores C , uma medida pareada ϖ , uma medida de centralidade χ , uma base de validação S_{val} , e um parâmetro p , responsável por forçar a avaliação de parte dos classificadores candidatos.

Depois de estabelecidos os parâmetros de entrada, estes têm a função de construir a rede (exceto por p). Este processo que é comum ao método de fusão é encontrado entre as linhas 1 a 12. É importante ressaltar que não é feita uma combinação ponderada de classificadores, mas que o processo de construção e análise da rede assemelha-se ao método de fusão *CBF*. Portanto, o objetivo desse trecho do algoritmo é montar a Rede de Classificadores e estimar a importância deles (Ψ) proveniente da pontuação sugerida pela medida de centralidade escolhida (χ). Cada classificador do *pool* C é adicionado a um conjunto de classificadores restante CR (linha 13). Este conjunto, CR , é responsável por armazenar apenas os classificadores que não fazem parte do subconjunto C' , que é construído ao longo das iterações e é o subconjunto final apresentado por *CBS*. Desta forma, $C = C' \cup CR$ e $C' \not\subset CR$. O subconjunto C' é iniciado vazio como pode ser visualizado na linha 14. Como a construção de C' inicia-se vazio e os classificadores são adicionados a cada iteração, este processo é inspirado pela técnica de *Forward Search*, vista na Seção 3.2.

Na linha 15 o conjunto CR é ordenado de acordo com o peso registrado em Ψ . O classificador CM de maior peso Ψ é adicionado ao subconjunto C' (linha 17) e removido do conjunto de classificadores candidatos CR (linha 18). O peso Ψ pode ser o obtido pelo classificador que mais complementa os demais se os relacionamentos da rede são representados por N_{10} ou é o classificador que mais diverge dos demais se os relacionamentos pareados são obtidos por *Estatística Q*, *Coeficiente de Correlação*, *Desacordo*, *Falta Dupla*

Algoritmo 7: CBS($C, \varpi, \chi, \tau, S_{val}, \varphi, p$)

Entrada: Conjunto de classificadores $C = \{c_1, c_2, \dots, c_T\}$
Entrada: Relação pareada (ϖ), onde $\varpi(c_i, c_j)$
Entrada: Medida de centralidade (χ)
Entrada: Base de validação (S_{val})
Entrada: Normalização (τ)
Entrada: Função de inversão (φ)
Entrada: Proporção de candidatos (p)

- 1 $V \leftarrow C$;
- 2 $E \leftarrow \varpi$;
- 3 **se** $\varphi = VERDADEIRO$ **então**
- 4 **para** $E_i \in E$ **faça**
- 5 $E_i = 1/E_i$;
- 6 **fim**
- 7 **se** $\tau = VERDADEIRO$ **então**
- 8 $E = normalizar(E)$;
- 9 **fim**
- 10 **fim**
- 11 Construir a rede $G(V, E)$;
- 12 $\Psi \leftarrow$ Estimar centralidade (χ, G);
- 13 $CR \leftarrow C$;
- 14 $C' \leftarrow \emptyset$;
- 15 ordenar CR pelo maior Ψ ;
- 16 $CM \leftarrow$ o classificador com maior Ψ presente em CR ;
- 17 $C' \leftarrow CM$;
- 18 $CR \leftarrow CR \setminus CM$;
- 19 $CC \leftarrow$ o classificador que mais complementa (ou que mais diverge de) CM ;
- 20 $C' \leftarrow CC$;
- 21 $CR \leftarrow CR \setminus CC$;
- 22 **repetir**
- 23 estimar acurácia $A_{C'}$ do conjunto C' ;
- 24 $CP \leftarrow$ uma parte dos classificadores em CR definida pela proporção p ;
- 25 **para** cada $c_i \in CP$ **faça**
- 26 estimar acurácia A_i do conjunto $c_i \cup C'$;
- 27 **fim**
- 28 $AM \leftarrow$ maior valor de acurácia;
- 29 $CM \leftarrow$ o classificador mais acurado;
- 30 **se** $AM \geq A_{C'}$ **então**
- 31 $C' \leftarrow CM \cup C'$;
- 32 $CR \leftarrow CR \setminus CM$;
- 33 **fim**
- 34 **até que** $AM \leq A_{C'}$;
- 35 **retorne** C' ;

ou a *Estatística Kappa*.

A linha 19 traz o classificador que mais complementa CM (CC), obtido pela análise das arestas que saem de CM e apontam para CC . Como mencionado anteriormente, também pode ser a diversidade entre eles, que é representada na rede por relações pareadas não direcionadas oriundas da estimativa de qualquer medida de diversidade pareada apresentada na Seção 2.4. Para as medidas de diversidade, é importante mencionar que quanto maior o peso maior a diversidade entre o par, como explicado no método de fusão CBF cujos parâmetros são encontrados na Tabela 4. Após identificar o classificador CC , ele é incluído no subgrupo C' (linha 20) e removido do grupo de classificadores candidatos CR (linha 21).

O processo de repetição apresentado nas linhas 22 e 34 determina como é adicionado o restante dos classificadores em CR ao subconjunto C' . A acurácia do conjunto C' é estimada na linha 23. Este passo serve para determinar se um classificador candidato CM será integrado ao conjunto futuramente.

Uma parte dos classificadores ponderados pelo score de centralidade (definido na linha 12) é escolhida para fazer parte do subconjunto CP . Este conjunto, que refere-se aos classificadores mais promissores ao subconjunto C' , representa os classificadores mais complementares ou mais diversos, obtidos do subconjunto restante CR , logo $CP \in CR$. A idéia de avaliar apenas parte dos classificadores restantes, é focar nos classificadores cuja contribuição para o conjunto é distinta, seja em termos de diversidade ou complementariedade. O conjunto CP tem seu tamanho definido pelo parâmetro $p = \{0, 1\}$, obtido em função da quantidade de classificadores que pertencem ao subconjunto CR (linha 24). Assim, apenas os classificadores com maior pontuação de centralidade são escolhidos para serem parte de CP . Cada candidato $c_i \in CP$ passa a ter a acurácia estimada (A_i) quando em união com o subconjunto C' (linha 26).

Após ordenar os classificadores pela acurácia A_i (linha 28), avalia-se a acurácia do subconjunto C' ($A_{C'}$) com a acurácia do candidato c_i (A_i) em união com o conjunto C' ($c_i \cup C'$) (linha 30). Se a maior acurácia encontrada, $AM \geq A_{C'}$, o classificador relacionado CM é adicionado ao subconjunto C' (linha 31) e removido do conjunto restante CR (linha 32). Caso contrário, o processo de repetição é finalizado (linha 34).

O Algoritmo 7 implementa na busca pelo subconjunto C' uma troca entre a centralidade do classificador baseada na importância de sua complementariedade para o conjunto e a acurácia que os mais complementares formam com o subconjunto já previamente formado, escolhendo pelo classificador mais acurado dentre os mais complementares. O

algoritmo DREP implementa um mecanismo semelhante, entretanto, enquanto este estima a complementariedade do classificador em relação ao subconjunto previamente selecionado, o CBS estima a complementariedade do classificador pela posição que este ocupa na rede.

4.4 Considerações finais

Neste capítulo foram apresentados dois métodos, um de combinação e outro de seleção. Estes, fazem uso de técnicas de Redes Complexas com o objetivo geral de elevar a acurácia de um SMC. O conceito de centralidade foi aplicado aos classificadores dispostos em rede, sugerindo que os classificadores podem ser ponderados.

A partir da ponderação dos classificadores, um processo de fusão estática foi desenvolvido. A abordagem CBF faz a análise dos classificadores dispostos em rede por meio da análise dos relacionamentos pareados entre eles. O método de fusão proposto é capaz de analisar redes ponderadas ou não ponderadas.

O mesmo conceito de centralidade, mediante a ponderação, permite que seja possível ordenar os classificadores. Isto é usado na abordagem CBS como um critério para a escolha dos classificadores mais promissores a fazerem parte de um subconjunto C' . Este subconjunto tem a função de aumentar ou manter a acurácia do conjunto C , sendo obtido pela remoção dos classificadores inicialmente gerados.

O próximo capítulo apresenta os experimentos pelos quais os métodos CBF e CBS são analisados. Pelos experimentos, algumas vantagens e limitações são apresentadas para cada um. Ambos os métodos também são comparadas com seus respectivos concorrentes encontrados na literatura.

5 Experimentos

Foram realizados experimentos em 30 problemas de classificação obtidos por diferentes repositório de dados, como UCI Machine Learning Repository (LICHMAN, 2013), KEEL (Knowledge Extraction based on Evolutionary Learning) Repository (ALCALÁ-FDEZ et al., 2011), Ludmila Kuncheva Collection (LKC) of Real Medical Data (KUNCHEVA, 2004), STATLOG project (KING; FENG; SUTHERLAND, 1995) e bases artificiais geradas com Matlab PRTools toolbox (DUIN et al., 2017). Estas bases de dados apresentam apenas atributos numéricos, além de um número diversificado de instâncias, classes, atributos e taxa de balanceamento de classes (proporção entre a classe majoritária em relação a minoritária). Além disso, há 20 problemas com apenas duas classes, enquanto os outros 10 apresentam 3 e 8 classes. Ecoli é o problema de classificação mais desbalanceado, de acordo com a taxa de balanceamento apresentada. A motivação por usar estes problemas tão diferentes entre si é para avaliar o desempenho e robustez do método de forma geral, assim como compará-lo a abordagens da literatura. Todos os problemas de classificação mencionados estão detalhados na Tabela 5.

O protocolo experimental foi baseado numa validação cruzada com 6 subconjuntos mutuamente exclusivos (*k-fold* com $k = 6$). Assim, $k - 1$ subconjuntos são utilizados para treinamento enquanto que 1 subconjunto é utilizado para teste. Após alternar entre os subconjuntos, é obtida a média entre eles, bem como o desvio-padrão. A validação cruzada tem a vantagem de não apresentar a sobreposição das instâncias de treino e teste. Além disso, cada k base de teste é exclusiva, evitando que o modelo testado sofra algum tipo de tendência devido uma má distribuição das instâncias em possíveis n repetições. Por fim, também é considerado como motivação a escolha de k -fold no lugar de n repetições, a qualidade na análise de menos testes, porém significativos, facilitando o entendimento dos modelos avaliados nos 30 problemas de classificação. O valor de $k = 6$ também já foi usado em outro trabalho recente na literatura de SMC (CAVALCANTI et al., 2016).

A proporção da base original S para treinamento S_{train} , validação S_{val} e teste S_{test} fica assim especificada: 3 subconjuntos para S_{train} (50% do tamanho de S), 2 subconjuntos para validação S_{val} ($\cong 32, 3\%$) e um ($\cong 16, 7\%$) para teste S_{test} . A base S_{val} é usada para estimar a acurácia individual e também para estimar as relações pareadas. A escolha por uma base de validação para realizar estas estimativas é para que seja evitada o *overfitting* (chamado também de sobre-ajuste), isto é, a hipótese é ajustada aos dados de treinamento

Tabela 5 – Características principais de cada problema de classificação. Número de Instâncias (# I). Número de atributos (# F). Número de Classes (# C). Taxa de Balanceamento (T.B.).

Base	# I	# A	# C	T.B.	Repositório
Australian	690	14	2	1.25	UCI
Banana	2000	2	2	1.00	PRTools
Blood	748	4	2	3.20	UCI
CTG	2126	21	3	9.40	UCI
Diabetes	766	8	2	1.86	UCI
Ecoli	336	7	8	71.50	UCI
Faults	1941	27	7	12.24	UCI
German	1000	24	2	2.33	STATLOG
Glass	214	9	6	8.44	UCI
Haberman	306	3	2	2.78	UCI
Heart	270	13	2	1.25	STATLOG
ILPD	583	10	6	2.49	UCI
Ionosphere	351	34	2	1.79	UCI
Laryngeal1	213	16	2	1.63	LKC
Laryngeal3	353	16	3	4.11	LKC
Lithuanian	2000	2	2	1.00	PRTools
Liver	345	6	2	1.38	UCI
Magic	19020	10	2	1.84	KEEL
Mammo	830	5	2	1.06	KEEL
Monk	432	6	2	1.12	KEEL
Phoneme	5404	5	2	2.41	ELENA
Segmentation	2310	19	7	1.00	UCI
Sonar	208	60	2	1.14	UCI
Thyroid	692	16	2	12.06	LKC
Vehicle	847	18	4	1.10	STATLOG
Vertebral	300	6	2	2.13	UCI
WBC	569	30	2	1.68	UCI
WDVG	5000	21	3	1.03	UCI
Weaning	302	17	2	1.00	LKC
Wine	178	13	3	1.48	UCI

na ocasião da criação do conjunto e, portanto, realizar estimativas utilizando a mesma base pode induzir a expectativas errôneas quando o mesmo modelo é aplicado em dados não utilizados para treinamento. Foi adotada a amostragem estratificada (stratified sampling) para a composição dos três subconjuntos de dados (S_{train} , S_{val} and S_{test}) a qual respeita a proporção de exemplos observados na base original S pertencentes a cada classe.

Bagging foi escolhido para a geração de conjunto de classificadores fracos e diversificados entre si, cujo tamanho é $T = 100$. Cada classificador é treinado em uma parte dos dados disponíveis (para treinamento) que corresponde a 66% da base S_{train} . É possível que para essa proporção haja classificadores mais acurados que quando treinados em 100% de S_{train} como observado em (MARTÍNEZ-MUÑOZ; SUÁREZ, 2010), além de ser utilizados em outros trabalhos como (KO et al., 2007).

O uso de Perceptron para treinar os classificadores é recomendado por (CRUZ; SABOURIN; CAVALCANTI, 2017), pois é capaz de detectar pequenas diferenças entre as porções de S_{train} usadas para constituir cada classificador. Neste trabalho foi usado o Perceptron de Erro Quadrado Mínimo (Perceptron with Minimum Square Error) (JUANG;

KATAGIRI, 1992). Este classificador é encontrado na Weka Datamining Tool (FRANK; HALL; WITTEN, 2016) (versão 3.9.1). Todos os outros parâmetros foram mantidos padrão, pois de acordo com Amancio et al. (AMANCIO et al., 2014) tem boa performance e evita a necessidade por encontrar uma parametrização ideal. Todos os classificadores tem acurácia maior que 50% (avaliada em S_{val}), portanto, estes classificadores têm melhor desempenho do que a simples escolha aleatória.

A ferramenta utilizada para criar e analisar a Rede de Classificadores é a Graphstream (DUTOT et al., 2007)(versão 1.3), portanto, a partir dela é que são estimadas as medidas de centralidade (apresentadas na Seção 2.1.3), as quais têm por finalidade pontuar os classificadores de acordo com a importância de cada um para o conjunto como um todo. Os experimentos foram separados em duas partes principais para detalhar os resultados do método de fusão na Seção 5.1 e os resultados do método de seleção em 5.2

5.1 Avaliação experimental de CBF

Para a avaliação do método de fusão CBF, dois conjuntos de experimentos foram realizados. Primeiro, na Seção 5.1.1 várias medidas pareadas combinadas com medidas de centralidade foram analisadas com o objetivo de verificar como estas medidas juntas influenciam na acurácia do conjunto. A partir desta análise espera-se que seja possível escolher a melhor combinação para o método quanto a medida de diversidade que compõe a rede e a centralidade a qual analisa os classificadores a partir destes relacionamentos. Estes parâmetros então passam a ser usados no segundo conjunto de experimentos, na Seção 5.1.3 que refere-se a comparação com 9 métodos da literatura. Para finalizar, a discussão acerca dos experimentos referentes CBF é apresentada na Seção 5.1.4.

5.1.1 Avaliação das medidas pareadas com as medidas de centralidade

O conjunto de experimentos apresentados nesta seção explica como foi comparada a relação pareada entre os classificadores do conjunto que compõe as arestas da rede, a medida de centralidade usada para avaliar a importância do classificador a partir dos relacionamentos e como isso afeta a acurácia do conjunto como um todo. A partir desta análise espera-se estimar qual é a combinação de relação/centralidade mais interessante para a acurácia do conjunto.

Foram avaliadas a combinação de 5 medidas de diversidade (*Falta Dupla - DF*, *Coefficiente de Correlação - CC*, *Desacordo - Dis*, *Estatística Q - QS*, e *Estatística Kappa*

- *KS*) com 4 medidas de centralidade (*Intermediação*, *Proximidade*, *Grau*, e *Autovetor*) resultando em 20 combinações entre relação/centralidade para a análise. Essas medidas de centralidade podem tanto ser usadas para avaliar relacionamentos simétricos quanto assimétricos. Neste trabalho, elas são usadas apenas para a avaliação dos relacionamentos simétricos. A relação assimétrica N_{10} é avaliada pela *Centralidade de Grau de Entrada*, pois esta é a mais simples medida de centralidade para relações direcionadas. Outras medidas para o relacionamento direcionado podem ser a *Centralidade de Intermediação* e a *Centralidade de Proximidade*, porém, estas demandam maior custo computacional devido o cálculo de todos os geodésicos da rede.

Os relacionamentos básicos entre classificadores (N_{00} , N_{01} , N_{10} , ou N_{11}) descritos na Tabela 1 são usados para estimar relacionamentos mais complexos como as medidas de diversidade *Falta Dupla* - *DF*, *Coeficiente de Correlação* - *CC*, *Desacordo* - *Dis*, *Estatística Q* - *QS*, e *Estatística Kappa* - *KS*. Entretanto, em alguns casos, ao menos um desses relacionamentos básicos não acontece em nenhum caso. Isso implica muitas vezes numa divisão por zero na estimativa das medidas de diversidade mencionadas. Para evitar isso, em *QS* e *CC*, sempre que detectado uma divisão por zero em decorrência da falta de um dos relacionamentos básicos, atribui-se $QS = 1.0$ ou $CC = 1.0$, como sugerido em (TSYMBAL; PECHENIZKIY; CUNNINGHAM, 2003). Geralmente, os relacionamentos que não ocorrem são os N_{10} e/ou o N_{01} , que exprimem diversidade, assim, é sugerido que *QS* e *CC* tenham valor 1.0 porque este valor refere-se a ausência de diversidade.

Todas as medidas de centralidade utilizadas consideram o peso da aresta para sua estimativa, assim, há um melhor aproveitamento das informações obtidas a partir do relacionamento entre pares de classificadores. Além disso, é importante lembrar que existem medidas de centralidade que julgam mais importantes um classificador com relacionamentos de peso baixo e há outras que consideram mais importante classificadores com relacionamentos de peso alto. Portanto, como o objetivo da maior parte dos relacionamentos pareados é expressar diversidade, a centralidade precisa fazer a correta leitura da rede quanto a esse quesito. Desta forma, a correta análise da rede depende de como a medida de centralidade interpreta o relacionamento representado na rede e o que este relacionamento tem a contribuir para a análise. A combinação destas medidas de diversidade pareadas com essas medidas de centralidade, implicam na criação de 21 novas medidas para a análise da importância dos classificadores. A correta análise da importância dos classificadores depende de como os relacionamentos expressam a diversidade, portanto, esse aspecto importante é detalhado na Seção 4.2.4, em especial na Tabela 2.

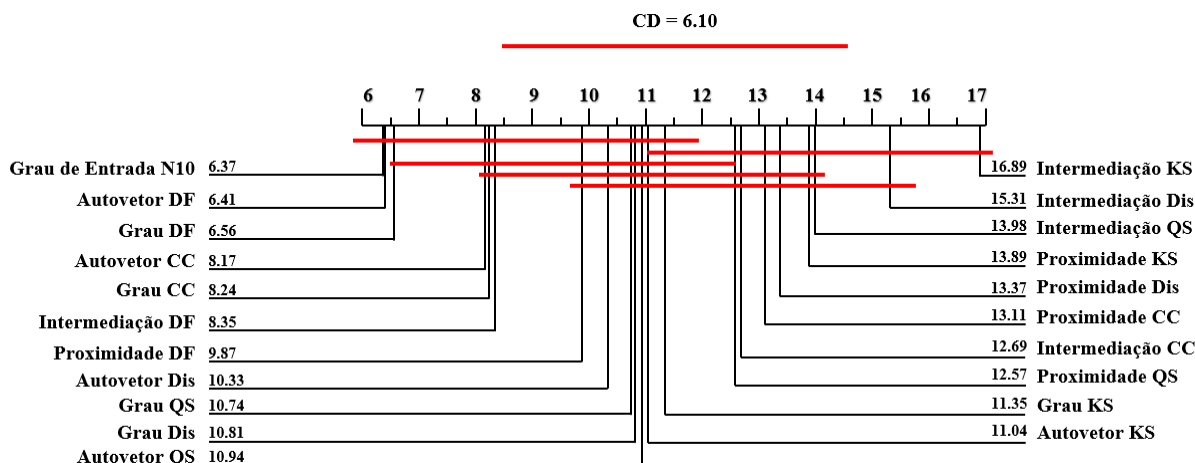


Figura 10 – Teste de Friedman e teste post-hoc Nemenyi para comparando a acurácia de CBF e a literatura.

Após a realização dos experimentos utilizando os 30 problemas de classificação, foram estimados o teste de Friedman e Nemenyi, os quais auxiliam na análise fornecendo dados estatísticos. Friedman estima a ordem classificatória dos resultados das diversas abordagens, neste caso, a combinação de relação/centralidade. Desta forma, o melhor resultado recebe 1, o segundo 2 e assim por diante. Para duas ou mais abordagens com o mesmo resultado, considera-se a média entre as duas posições mais próximas na classificação. Por exemplo, supondo que dois resultados poderiam ocupar respectivamente a posição 3 ou 4, pela situação gerada pelo empate, opta-se por estimar a média entre 3 e 4, logo, a classificação desses resultados fica 3.5. Após realizar a classificação de cada uma das 30 bases usando *k-fold* com $k = 6$, é obtida a média por problema e em seguida estima-se a média entre todos os 30 problemas. Como o índice mais baixo refere-se à melhor solução, a média mais baixa entre as abordagens sugere que uma solução está frequentemente entre as melhores observadas.

A Figura 10 mostra o posicionamento das diferentes medidas de centralidades e das medidas de diversidades (Estatística Q (QS), Coeficiente de Correlação (CC), Desacordo (Dis), Falta Dupla (DF), Estatística Kappa(KS) e o relacionamento pareado assimétrico N_{10}). Esta classificação sugere que a medida de centralidade *Grau de Saída* estimada numa rede cujas relações pareadas representam N_{10} é possivelmente uma das melhores combinações entre relação e centralidade, sendo referido nas próximas seções como (*CBF: Grau de Saída N_{10}*). Esta combinação de parâmetros sugere que ponderar os classificadores pela capacidade de complementar a decisão de seus pares é a melhor estratégia dentre aquelas comparadas para a fusão de classificadores. Com uma posição média muito próxima de *CBF: Grau de Saída N_{10}* estão *CBF: Autovetor DF* e *CBF: Grau DF*. As duas medidas

de centralidade, que foram estimadas em uma Rede de Classificadores que usa Falta Dupla para a representação dos relacionamentos, diferem-se entre si quanto a abrangência dos vértices avaliados, a *Centralidade de Grau* apenas considera os vizinhos diretos de cada vértice enquanto a *Centralidade de Autovetor* considera os vizinhos diretos e os vizinhos diretos desses vizinhos. A relação pareada que possui melhor posição é a DF, o que sugere que pode ser mais interessante mapear os classificadores que mais evitam o erro mútuo com seus pares que apenas avaliar a diversidade. Quanto à centralidade, a que menos se posicionou entre as mais interessantes soluções foi a *Centralidade de Intermediação*.

Para entender o motivo desta medida de centralidade ganhar este destaque negativo, foram conferidas as pontuações associadas a cada centralidade pela análise de cada tipo de relacionamento proposto. A *Centralidade de Intermediação* apresentou pontuações entre os classificadores frequentemente iguais e, portanto, neste caso, se comportaram como o voto majoritário caso o peso do classificador $\Psi_{c_i} > 0$. Nos casos em que havia ao menos um classificador com peso $\Psi_{c_i} \neq 0$ o conjunto votou pela classe que recebeu o maior peso a partir dos classificadores que não apresentavam peso nulo associado ao voto. Antes de apresentar a razão pela qual a *Centralidade de Intermediação* atribui peso 0 (zero) aos classificadores, é importante lembrar que esta é uma medida que conta a quantidade de vezes que o classificador está presente nos geodésicos. Numa rede completa, se os relacionamentos entre os classificadores possuírem pouca variação entre si a ponto do geodésico não envolver mais de dois classificadores (vértice inicial e o vértice final), estimar a presença do vértice nos geodésicos não é possível.

Considerando as medidas pareadas, DF foi a medida avaliada que mais esteve relacionada com a acurácia do conjunto. Esta medida pareada avalia a quantidade de erros simultâneos do par de classificadores. Isso é útil para a combinação de classificadores, pois elementos que juntos não acertam, não podem aumentar a acurácia do conjunto. Pelo contrário, evitar tais classificadores pode ajudar o conjunto a ter uma taxa de reconhecimento melhor. Portanto, DF se posicionou entre as melhores soluções se comparada as demais medidas simétricas. As medidas de centralidade que se mostraram mais interessantes foram a *Centralidade de Autovetor* e a *Centralidade de Grau*, como é possível observar na Figura 10, ocupando as primeiras posições tanto pelo relacionamento com DF quanto CC. Estas informações também sugerem que ao ponderar-se os classificadores, apenas pela divergência entre si, talvez isso não seja tão interessante quanto explorar e usar a favor as relações que induzem ao erro como DF ou relações de complementariedade no caso da relação N_{10} .

Embora a centralidade de *Grau de Entrada* estimada numa rede formada por relacionamentos N_{10} (*CBF: Grau de Entrada N_{10}*) ser estatisticamente diferente e ter uma taxa de reconhecimento mais atrativa que as demais avaliadas a razão deste critério de ponderação funcionar precisa ser investigada. Na Tabela 6 está apresentado o peso atribuído pela nova abordagem *CBF: Grau de Entrada N_{10}* . O peso do classificador foi comparado com o peso atribuído pela acurácia individual e o peso do relacionamento em média (obtido pelo peso dividido pelo número de vizinhos que apontam para o vértice, logo $T - 1 = 99$). O relacionamento médio é obtido pelo relacionamento do vértice identificado por *id* e os vizinhos que apontam para ele. O classificador 52 apresenta o maior peso além de ser um dos mais acurados (estimativa em S_{val}). No final da tabela, é apresentado o teste de correlação de Pearson. A correlação entre o peso do classificador e a medida pareada N_{10} é alto (como esperado) e também tem uma correlação significativamente positiva do peso do classificador com a acurácia individual (0.68). Note que ponderar o classificador pela acurácia individual e por *CBF* nesta rede não é a mesma coisa, mas possuem significativa correlação positiva.

A média dos relacionamentos de um classificador é a média de seus relacionamentos diretos (neste caso somente os relacionamentos que apontam para o vértice são válidos), e portanto, para N_{10} quanto maior essa média, maior é o complemento entre os classificadores. O relacionamento médio do conjunto é a média de todos os relacionamentos. Logo, um classificador cujo peso é maior que a média dos relacionamentos é um classificador interessante, pois o peso é computado pela sua contribuição total aos vizinhos diretos (que recorrem a ele quando estão incorretos). O classificador 3 apresenta baixo peso e também a pior acurácia. Seu relacionamento médio também é menor que a média do conjunto. Estas observações extraídas da tabela mostraram que o classificador mais importante segundo o método é um classificador acurado, mas mais importante que isso, é aquele que complementa a decisão de seus pares.

Para facilitar a visualização do relacionamento entre a acurácia individual dos classificadores e a média dos relacionamentos que ele possui com os demais é apresentada a Figura 11. A linha em vermelho representa a tendência linear simples. A tendência indica que quanto mais acurado um classificador, mais complementar ele é em relação a seus pares. Entretanto, o valor de R-Quadrado sugere que a linha não é confiável, pois o valor não está próximo de 1.0 , o qual indica a confiança máxima na tendência sinalizada pela linha. O valor de R-Quadrado sugere que há uma tendência entre os dados mas que não há uma correlação direta entre a acurácia individual dos classificadores e o relacionamento

Tabela 6 – Um caso de sucesso com a *Centralidade de Grau de Entrada* estimada em relacionamentos pareados N_{10} . A coluna (rel média) representa a média do peso dos relacionamentos considerando apenas os vizinhos que apontam para o vértice que é identificado por *id*.

id	peso	acurácia	rel média	id	peso	acurácia	rel média
0	5858	67.23	59.17	50	5638	66.00	56.95
1	6331	66.46	63.95	51	5411	66.77	54.66
2	6044	67.23	61.05	52	7628	68.93	77.05
3	4361	61.67	44.05	53	5931	67.08	59.91
4	6343	67.54	64.07	54	5133	66.62	51.85
5	5460	64.45	55.15	55	5912	65.38	59.72
6	6532	68.16	65.98	56	5981	66.62	60.41
7	5205	62.29	52.58	57	5530	65.53	55.86
8	5319	65.07	53.73	58	6040	68.32	61.01
9	5916	67.54	59.76	59	5658	66.31	57.15
10	5129	64.61	51.81	60	5292	65.53	53.45
11	5064	65.69	51.15	61	5064	64.45	51.15
12	5689	66.46	57.46	62	5499	66.77	55.55
13	6737	68.01	68.05	63	6231	65.53	62.94
14	5420	65.53	54.75	64	6677	68.32	67.44
15	5480	63.37	55.35	65	5599	66.92	56.56
16	6793	66.92	68.62	66	5568	66.92	56.24
17	5968	67.70	60.28	67	6748	67.39	68.16
18	5326	65.53	53.80	68	4912	63.68	49.62
19	6835	67.08	69.04	69	5202	65.07	52.55
20	5573	65.84	56.29	70	6419	67.39	64.84
21	5163	66.15	52.15	71	6179	68.32	62.41
22	6073	66.46	61.34	72	6176	67.85	62.38
23	5115	66.15	51.67	73	6142	66.15	62.04
24	5925	68.01	59.85	74	7173	68.16	72.45
25	6949	67.54	70.19	75	6558	69.71	66.24
26	6250	67.39	63.13	76	5493	63.52	55.48
27	5460	66.15	55.15	77	6659	69.55	67.26
28	6207	67.08	62.70	78	5779	66.62	58.37
29	5668	65.07	57.25	79	5587	67.23	56.43
30	6218	67.70	62.81	80	6568	66.46	66.34
31	5757	65.53	58.15	81	5555	66.77	56.11
32	6015	67.54	60.76	82	5549	60.90	56.05
33	4857	65.69	49.06	83	5866	65.22	59.25
34	5958	65.84	60.18	84	6106	66.15	61.68
35	5293	65.07	53.46	85	5812	67.23	58.71
36	5235	65.38	52.88	86	6061	65.22	61.22
37	5047	65.69	50.98	87	6046	67.23	61.07
38	5696	66.00	57.54	88	5468	67.08	55.23
39	5985	66.00	60.45	89	5970	67.08	60.30
40	5658	64.91	57.15	90	5969	67.39	60.29
41	5988	67.39	60.48	91	5252	63.83	53.05
42	5735	67.23	57.93	92	6099	66.92	61.61
43	5336	65.22	53.90	93	5899	65.07	59.59
44	5885	68.47	59.44	94	5511	65.69	55.67
45	5999	65.07	60.60	95	6436	66.62	65.01
46	5087	64.14	51.38	96	6248	67.70	63.11
47	5323	63.37	53.77	97	5717	65.38	57.75
48	5879	68.47	59.38	98	5048	61.05	50.99
49	5704	66.31	57.62	99	6461	68.62	65.26
Pearson peso/acurácia							0.68
Pearson peso/relacionamento							1
Pearson acurácia/relacionamento							0.68
Média dos relacionamentos do ensemble							58.82

médio que ele possui com os pares, neste caso, a relação de complementariedade N_{10} .

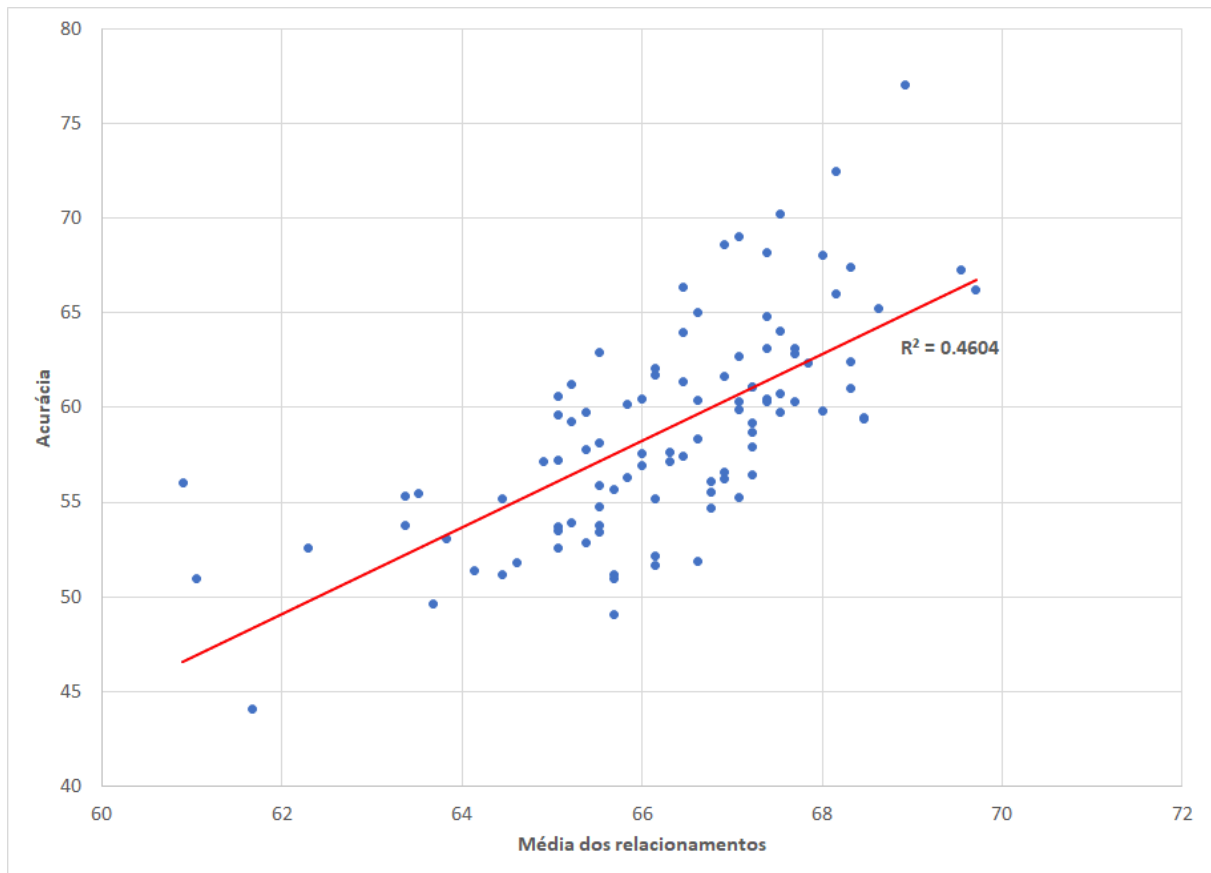


Figura 11 – Correlação entre a acurácia individual dos classificadores e o peso médio dos relacionamentos.

5.1.1.1 Avaliação das centralidades numa rede ponderada e simplificada

Nesta seção, o método CBF é usado para ponderar os classificadores numa rede não completa. A rede, inicialmente completa, foi simplificada pelo algoritmo Poda Ingênuo Simplificada apresentado na Seção 2.1.4. Este algoritmo, difere do original (*Naive Approach Algorithm*) pelo critério de parada. O original segue a poda dos relacionamentos até atingir a proporção informada pelo usuário, ignorando a poda de relacionamentos ponte no processo, isto é, aqueles cuja remoção aumentam o número de componentes. O simplificado, entretanto, pára ao identificar um relacionamento ponte. O critério de parada adotado neste algoritmo apresenta a vantagem de não requerer do usuário a proporção de relacionamentos a serem removidos.

A simplificação de rede é motivada geralmente pela facilidade de visualização e análise. Entretanto, ao remover arestas e/ou vértices, há uma perda de informação que demanda cuidado. O tipo de remoção deste algoritmo é a remoção de arestas, que são

ordenadas pela diversidade. Desta forma, a rede simplificada tem menos arestas e apenas os relacionamentos de mais alta diversidade entre os pares permanecem. Além das motivações já mencionadas (visualização e análise), a simplificação da rede permite que a *Centralidade de Intermediação* seja estimada. Ao simplificar uma rede completa, aumenta-se o número de intermediários nos geodésicos. Isto acontece principalmente se a rede completa apresenta valores de arestas com pouca diferença.

Um dos problemas observados pelo experimento com a rede completa foi de que as arestas possuem, em geral, uma variação de valores pequena. Isto dificulta a estimativa dos classificadores que intermediam os relacionamentos entre os pares. O uso de uma rede não completa facilita a estimativa de todas as medidas de centralidade, pois, são menos vértices e/ou arestas a serem analisadas.

A Figura 12 mostra o Teste de Friedman e Nemenyi para uma rede inicialmente completa, simplificada pelo algoritmo de Poda Ingênua Simplificada. Percebe-se que a *Centralidade de Intermediação*, inicialmente uma das piores centralidades avaliadas, passa a ser uma opção interessante, mas abaixo da simples *Centralidade de Grau*.

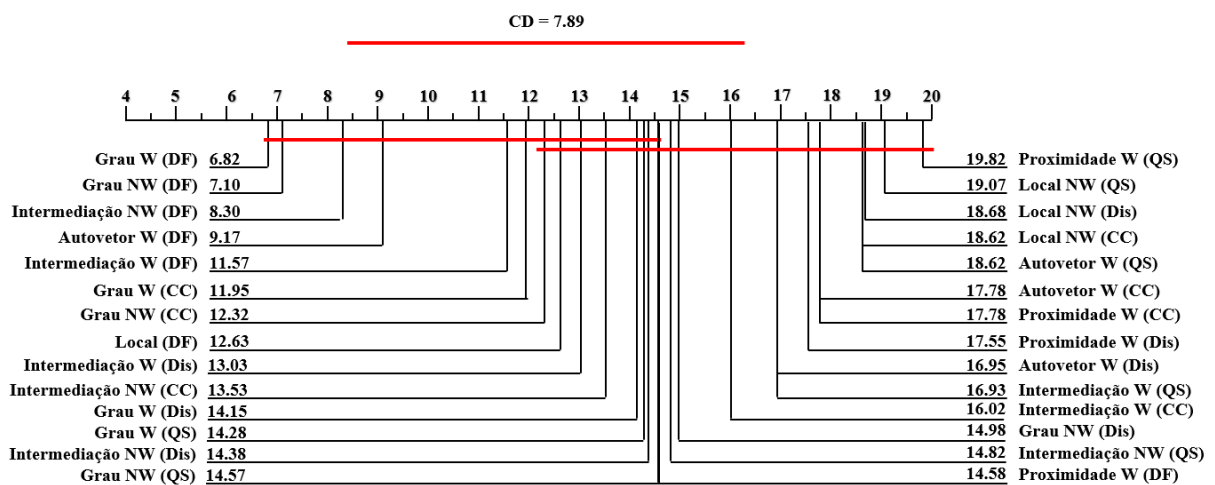


Figura 12 – Teste de Friedman e teste post-hoc Nemenyi comparando a acurácia de CBF numa rede simplificada e a literatura.

Neste experimento, somente as medidas pareadas simétricas formaram o conjunto de relacionamentos da rede que foi simplificada. Apesar da simplificação de rede ser empregada por diversos autores, é importante considerar que há sempre o dilema da perda de informação para facilitar a visualização e análise da rede (TRAWINSKI; CORDON, 2016; ZHOU, 2012; NEWMAN, 2010).

5.1.2 Relação entre a acurácia dos classificadores e a centralidade

A acurácia individual dos classificadores é utilizada comumente tanto em métodos de fusão quanto de seleção. Diante deste cenário, é importante avaliar qual é a relação entre a acurácia individual dos membros do *pool* e a pontuação referente à posição do classificador na rede (centralidade). A Figura 13 apresenta como estes dois atributos se relacionam. Note que a acurácia do classificador não foi normalizada, mas, sim, a importância do classificador (obtida pela análise da centralidade de rede). Os exemplos contidos nesta figura referem-se a análise de um experimento realizado na base de dados P2Problem. Assim, este mesmo caso, quando analisado por diferentes medidas de centralidade, apresenta diferentes perspectivas quanto a importância do classificador. É possível notar isso pela mudança do índice de centralidade enquanto que a acurácia do classificador é mantida.

Todos os 100 classificadores usados no protocolo experimental estão contidos neste exemplo. Portanto, há em vários casos a sobreposição da pontuação referente a centralidade do classificador, bem como a acurácia dos mesmos. Considerando a centralidade dos classificadores em função dos relacionamentos de diversidade e a acurácia individual, algumas destas situações podem acontecer:

- Um par de classificadores podem ter acurácia e centralidade diferentes entre si;
- Um par de classificadores podem ter acurácia igual, mas apresentar centralidade diferente;

O primeiro caso acontece com a maioria dos classificadores representados na Figura 13. Entretanto, há classificadores que tem tanto a acurácia quanto a centralidade (estimada em relações de diversidade) iguais. Neste caso, há forte similaridade entre os classificadores. O segundo caso acontece comparando os pontos na horizontal, pois apesar da mesma acurácia, possuem pontuação de centralidade diferentes. Portanto, isto sugere que um classificador, contribui mais para a diversidade do *ensemble* que o outro. É importante lembrar que a centralidade do classificador, foi normalizada entre 0 e 1 apenas para ilustração, pois, como já foi mencionado no capítulo de proposição de método, classificadores cujo peso no voto correspondem a zero não emitem opinião no processo classificatório.

A correlação entre a acurácia individual dos classificadores e a centralidade deles é significativamente positiva para as seguintes centralidades / relacionamentos: *Grau de Entrada* N_{10} , *Proximidade DF*, *Autovetor DF* e *Grau DF*. As demais medidas de diversidade estão menos correlacionadas com a acurácia individual dos classificadores.

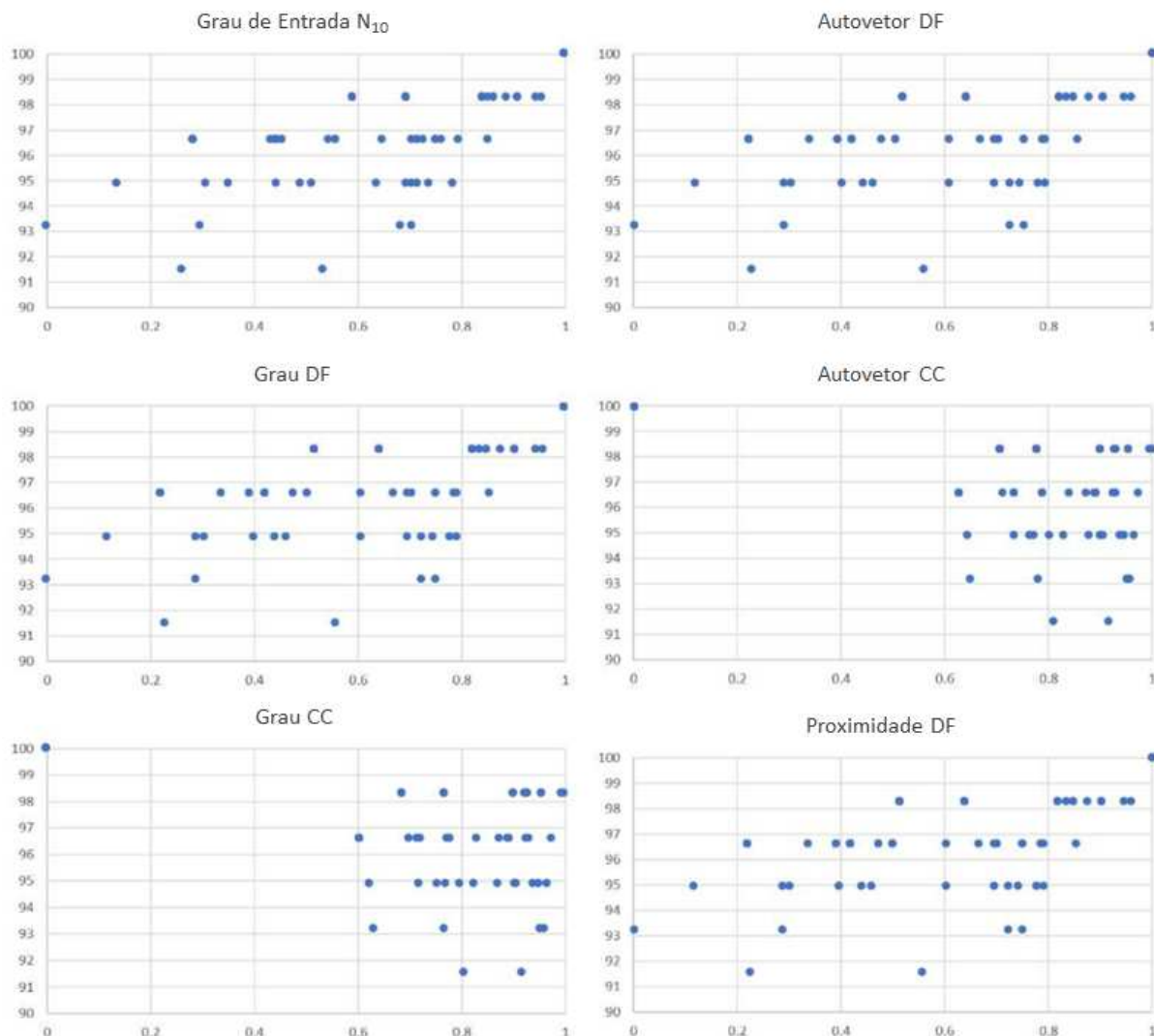


Figura 13 – Relação entre a importância (normalizada) do classificador (eixo x) e a acurácia (eixo y) para a base P2Problem com geração de pool por Bagging.

Portanto, estas medidas não desempenham tão bem por não indicar a complementariedade das decisões do conjunto ou a redução do erro como as medidas N_{10} e Falta Dupla (DF) sugerem respectivamente. Isto sugere que para relacionar a centralidade dos classificadores com a acurácia individual deles, é importante que o relacionamento expresso na rede de alguma forma indique acerto ou erro, e não apenas a diversidade de opiniões.

Continuando a avaliar as centralidades estimadas na rede DF e N_{10} , é possível perceber pequenas mudanças na ordem dos classificadores apresentados na Tabela 7. O classificador mais acurado (0) possui neste caso a mesma posição quanto à centralidade de Proximidade, Grau, Autovetor e Grau de Entrada. Nesta Tabela 7, percebe-se que a partir da diretriz para a construção da rede de classificadores apresentada na seção de métodos, a estimativa da centralidade apresenta valores diferentes para cada classificador, mesmo

que a ordem seja semelhante.

Tabela 7 – Classificadores ordenados pela centralidade. A coluna (id) refere-se a identificação do classificador. A coluna (valor) indica o peso atribuído pela medida de centralidade e na coluna (valor N.) é representado o valor de centralidade normalizado. Os valores foram obtidos pela análise da base P2Problem. Foram gerados 20 classificadores para este exemplo.

	Acurácia			Proximidade DF			Grau DF			Autovetor DF			Grau de Entrada N10		
	id	valor	valor N.	id	valor	valor N.	id	valor	valor N.	id	valor	valor N.	id	valor	valor N.
	6	69.28	1.00	8	0.041545	1.00	8	32.93	1.00	8	0.226771	1.00	8	369	1.00
	13	69.28	1.00	14	0.041459	0.95	14	32.88	0.95	14	0.226447	0.95	14	358	0.94
	8	68.67	0.97	3	0.041356	0.88	3	32.82	0.89	3	0.226062	0.89	3	348	0.88
	14	68.67	0.97	4	0.041305	0.85	4	32.79	0.86	4	0.225865	0.86	6	337	0.82
	11	67.47	0.90	6	0.041169	0.77	6	32.71	0.78	6	0.225346	0.78	13	337	0.82
	15	67.47	0.90	13	0.041169	0.77	13	32.71	0.78	13	0.225346	0.78	4	325	0.76
	3	66.87	0.87	9	0.041118	0.74	9	32.68	0.75	9	0.22513	0.74	9	307	0.66
	17	65.66	0.80	11	0.0409	0.61	11	32.55	0.62	11	0.224292	0.61	11	304	0.64
	5	64.46	0.73	15	0.0409	0.61	15	32.55	0.62	15	0.224292	0.61	15	304	0.64
	4	59.64	0.47	12	0.040866	0.59	12	32.53	0.60	12	0.22415	0.59	12	285	0.54
	18	59.64	0.47	17	0.04075	0.52	17	32.46	0.53	19	0.223766	0.53	17	283	0.52
	0	59.04	0.43	19	0.04075	0.52	19	32.46	0.53	17	0.223701	0.52	19	267	0.44
	2	59.04	0.43	5	0.040519	0.37	5	32.32	0.38	5	0.222814	0.38	5	260	0.40
	12	58.43	0.40	10	0.040258	0.22	10	32.16	0.22	10	0.221794	0.23	10	222	0.19
	1	57.83	0.37	2	0.040225	0.20	2	32.14	0.20	2	0.221662	0.20	2	219	0.17
	10	57.83	0.37	7	0.040161	0.16	7	32.1	0.16	7	0.221367	0.16	0	208	0.11
	16	57.23	0.33	0	0.040112	0.13	0	32.07	0.13	0	0.221199	0.13	18	206	0.10
	19	57.23	0.33	18	0.040064	0.10	18	32.04	0.10	18	0.220991	0.10	7	201	0.07
	9	56.63	0.30	1	0.039968	0.04	1	31.98	0.04	1	0.220608	0.04	1	191	0.02
	7	51.2	0.00	16	0.039904	0.00	16	31.94	0.00	16	0.220346	0.00	16	188	0.00
Média		62.08	0.60		0.0407	0.50		32.44	0.51		0.2236	0.51		275.95	0.49
Desvio-Padrão		5.32	0.29		0.0005	0.32		0.32	0.32		0.0020	0.32		59.02	0.33

A Tabela 7 mostra os valores originais atribuídos pelas medidas de centralidade e ao lado os valores de centralidade normalizados (apresentado na coluna valor N.). Percebe-se que o peso do classificador é diferente entre as medidas de centralidade mesmo que normalizado. Além disso, a média e o desvio-padrão normalizados indicam que o peso atribuído pelas medidas de centralidade tem uma variação menor que o peso atribuído pela acurácia individual dos classificadores.

A Figura 14 apresenta as redes detalhadas na Tabela 7, destacando as arestas de maior peso como mais espessas e mais escuras. A rede direcionada tem o valor dos relacionamentos definido por N_{10} enquanto que a não direcionada tem o peso definido pela medida de diversidade Falta Dupla. Esta última, não demanda transformação no valor do peso dos relacionamentos para o cálculo da *Centralidade de Proximidade*, pois quanto menor o valor da Falta Dupla, maior é o peso dado pela centralidade. Entretanto, para a avaliação da *Centralidade de Grau* e da *Centralidade de Autovetor*, o peso dos relacionamentos deve seguir as recomendações sugeridas no capítulo de proposição de método. Essas centralidades atribuem peso maior a relacionamentos de maior peso. Entretanto, no caso da Falta Dupla, quanto maior o erro do par dos classificadores maior é o peso e, por consequência, maior é o peso destas medidas de centralidade. Para atribuir uma pontuação de centralidade maior para classificadores que juntos erram menos, o peso sugerido pela Falta Dupla precisa ser adaptado.

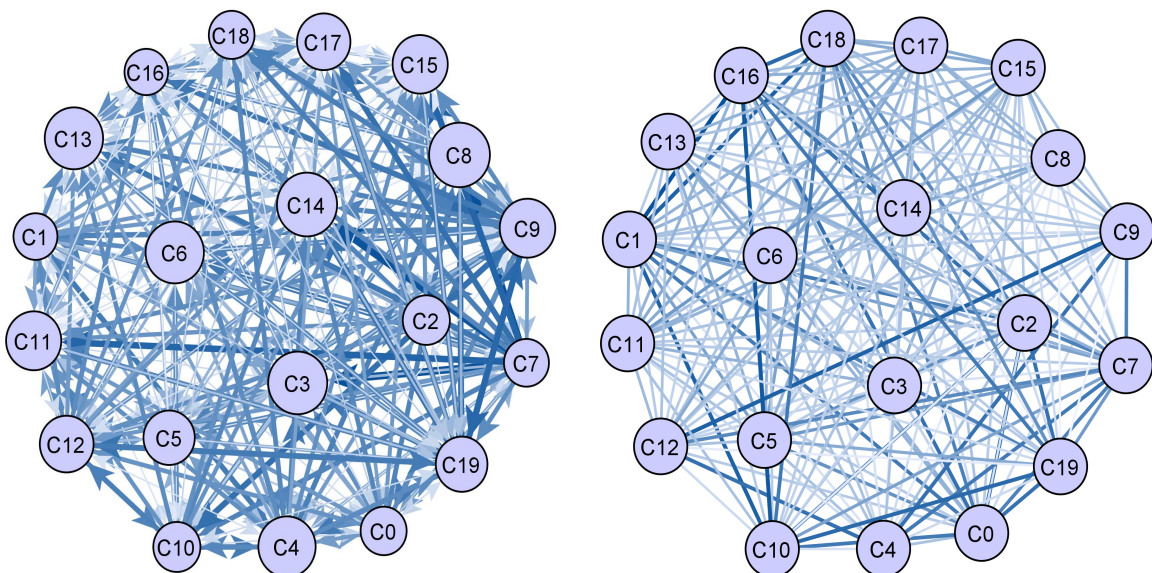


Figura 14 – Redes formadas por 20 classificadores gerados pelo método Bagging para a base P2Problem.

Nesta seção, foram destacadas as diferenças entre as medidas de centralidade, as

quais analisaram relacionamentos formados por medidas de diversidade pareadas. Nota-se que algumas medidas de centralidade apresentam maior desvio-padrão, influenciando diretamente o peso associado ao voto dos classificadores. Na próxima seção, a abordagem CBF é comparada com abordagens da literatura.

5.1.3 Comparação com métodos de fusão da literatura

Nesta seção o desempenho quanto à capacidade de reconhecimento de padrões de *CBF: Grau de Entrada* N_{10} - o qual computa a importância dos classificadores em uma rede de classificadores cujo relacionamento é representado por N_{10} e a análise da pontuação é dada pela medida de centralidade Grau de Entrada. A rede N_{10} usada não foi simplificada, portanto, a rede de classificadores apresenta-se como uma rede completa. A escolha de *CBF: Grau de Entrada* N_{10} é devido a análise apresentada na Seção 5.1.1. Foram avaliados 9 métodos da literatura: a) Voto Majoritário (MV); b) Voto Majoritário ponderado por Acurácia (WMV); c) Ponderação por Performance (PW); d) Voto Majoritário Ponderado de Kuncheva (KWMV); e) Combinação Bayesiana (BC); f) Valor de Poder (PV); g) Voto Ponderado por Redimensionamento (RSWV); h) Melhor-Pior Voto Ponderado Quadrático (QBWWV); e i) Melhor-Pior Voto Ponderado (BWWV). Todas estas abordagens estão descritas na Seção 3.1.

Tabela 8 – Acurácia média e o desvio-padrão de cada abordagem avaliada baseada em validação cruzada 6-fold. WS significa teste dos postos sinalizados de Wilcoxon, portanto, os valores representam o valor de p, e + é para resultados significativos. Os melhores resultados estão em negrito. # Best é o número de melhores resultados obtidos em todos os problemas de classificação.

	MV	WMV	PV	RSWV	QBWWV	BWWV	PW	KWMV	BC	<i>CBF: Grau Ent. N₁₀</i>
Australian	87.83 ±3.33	87.83 ±3.33	87.54±3.43	69.86±6.16	78.12±2.63	85.65±3.12	87.54±3.64	87.68±3.45	87.54±3.45	87.25±2.87
Banana	85.10±2.35	85.10±2.36	85.10±2.36	84.40±1.61	84.30±1.94	84.75±1.77	85.00±2.32	85.10±2.36	85.15 ±2.36	84.80±1.99
Blood	78.07±1.21	78.07±1.21	77.94±1.29	76.34±0.53	75.66±2.39	78.88 ±1.71	77.94±1.29	77.94±1.29	78.07±1.29	77.81±1.38
CTG	89.56±0.66	89.65±0.69	89.60±0.69	83.21±0.80	87.82±1.25	88.76±0.92	89.60±0.92	89.60±0.69	89.27±0.69	89.75 ±0.82
Diabetes	76.64±5.17	76.77±5.37	76.90±5.39	69.20±2.39	71.42±3.34	76.24±3.85	77.03±5.17	76.90±5.39	77.03±5.39	77.16 ±5.09
Ecoli	86.01±5.50	86.01±5.50	86.61 ±5.13	75.30±0.67	80.66±4.18	85.12±3.95	86.31±4.69	86.31±5.32	86.61±5.32	86.31±4.08
Faults	70.84±0.96	70.84±1.02	70.90±1.03	49.66±4.60	64.24±2.10	66.87±1.92	70.84±1.02	70.84±1.10	70.07±1.10	71.15 ±0.99
German	75.50±2.51	75.60±2.59	76.00 ±2.47	70.20±0.48	67.29±5.01	73.20±3.84	75.90±2.58	75.70±2.51	75.40±2.51	75.90±2.09
Glass	61.18±5.46	61.17±6.83	61.19±6.50	45.29±4.85	54.64±4.37	60.73±6.52	61.19±6.50	62.12±7.00	62.09±7.00	63.06 ±7.81
Haberman	74.18±4.44	74.18±4.44	73.86±4.76	73.20±0.92	75.16 ±4.03	72.88±6.64	74.18±4.44	74.18±4.44	73.86±4.44	73.20±6.37
Heart	83.70 ±6.75	83.33±6.12	83.33±5.70	61.11±3.57	72.22±5.25	79.26±4.74	83.70 ±6.24	83.70 ±6.24	83.70 ±6.24	82.96±5.54
ILPD	71.36±1.98	71.18±2.04	71.18±2.86	71.35±0.38	64.16±3.43	70.49±3.24	71.36±2.82	71.53 ±2.48	71.53±2.48	71.19±2.84
Ionosphere	85.76±3.36	85.76±3.36	85.76±3.36	65.24±0.86	77.50±4.53	84.05±4.59	85.76±3.36	85.76±3.36	85.76±3.36	86.04 ±3.84
Laryngeal1	80.30±5.54	79.36±5.66	80.30±5.02	57.28±4.92	72.38±9.36	78.39±4.86	79.83±4.89	79.83±4.89	78.90±4.89	81.24 ±4.96
Laryngeal3	73.95±3.12	74.24 ±3.50	74.24±3.36	62.60±4.43	63.16±7.47	70.54±2.87	74.24 ±3.50	73.96±3.91	73.38±3.91	73.10±3.53
Lithuanian	83.10±2.10	83.10±2.10	83.10±2.10	82.85±1.22	82.30±1.87	82.75±2.01	83.20 ±2.19	83.10±2.10	82.80±2.10	83.10±1.79
Liver	68.70±3.77	68.12±4.19	68.69±4.03	54.24±5.90	59.45±6.42	70.47 ±5.49	68.99±4.06	68.70±4.23	68.70±4.23	69.27±4.70
Magic	79.29±0.49	79.28±0.45	79.29±0.44	77.63±0.32	78.44±0.42	79.57 ±0.64	79.38±0.56	79.28±0.45	79.27±0.45	79.42±0.68
Mammo	83.37±2.31	83.49±2.36	84.10±2.24	77.71±1.84	80.12±2.82	83.86±2.49	84.10±2.36	83.73±2.39	83.86±2.39	84.58 ±1.75
Monk	81.48±2.96	82.41±2.49	83.33±2.89	75.00±4.17	79.17±3.30	88.43 ±2.74	83.10±2.82	82.87±2.85	82.41±2.85	86.11±3.50
Phoneme	77.17±0.86	77.18±0.89	77.37±0.94	75.42±1.31	73.43±1.21	77.81 ±0.99	77.54±0.97	77.22±0.92	77.24±0.92	77.59±1.14
Segmentation	92.64±1.69	92.60±1.67	92.56±1.64	87.36±1.23	90.26±1.51	92.99 ±0.88	92.82±1.53	92.60±1.67	92.69±1.67	92.82±1.30
Sonar	79.34±6.67	79.34±6.67	79.37±7.92	59.64±4.17	74.58±7.13	74.50±7.37	79.36±7.57	79.83±6.54	78.85±6.54	79.85 ±6.94
Thyroid	96.39±1.17	96.39±1.17	96.39±1.17	86.71±1.85	93.93±2.61	96.38±2.33	96.53±1.13	96.39±1.17	96.24±1.17	96.67 ±0.93
Vehicle	76.95±2.15	76.83±1.77	76.95±1.95	65.84±2.26	73.17±2.19	75.30±2.29	76.95±1.95	76.83±1.77	76.48±1.77	77.31 ±3.06
Vertebral	87.00 ±4.12	87.00 ±4.12	86.67±3.94	78.33±2.69	82.33±3.35	85.67±3.35	86.67±3.94	86.67±3.94	86.00±3.94	86.33±4.96
WBC	96.67±1.65	96.49±1.79	96.49±1.79	89.99±3.08	93.86±1.95	95.96±1.12	96.67±1.65	96.49±1.79	96.14±1.79	97.02 ±1.54
WDVG	86.32±0.74	86.30±0.73	86.30±0.73	81.60±1.21	84.14±1.46	85.82±0.92	86.36±0.68	86.30±0.73	86.38 ±0.73	86.38 ±0.57
Weaning	82.10±3.32	82.10±3.32	82.10±3.73	57.60±3.53	69.52±7.48	79.77±5.04	81.77±3.96	81.77±3.22	81.77±3.22	82.43 ±3.23
Wine	98.85 ±1.63	98.85 ±1.63	98.85 ±1.63	85.98±4.42	91.57±3.20	94.96±1.63	98.85 ±1.63	96.07±2.28	98.85 ±2.28	98.85 ±1.63
# Best	4	4	3	0	1	6	4	2	2	14
WS	0.09	0.05	0.12	+0.00	+0.00	+0.00	0.18	0.08	+0.01	n/a

A Tabela 8 mostra a acurácia média e o desvio-padrão correspondente do método proposto em comparação com 9 métodos de fusão da literatura. Como pode ser notado, a abordagem CBF foi capaz de alcançar o melhor resultado em 14 de 30 problemas, enquanto não registra nenhum pior resultado. Destas 14 bases, 10 possuem a quantidade de atributos superior a 14, o que sugere que CBF é melhor adaptado para bases com maior quantidade de atributos.

Nesta tabela percebe-se que o desempenho das diferentes abordagens de combinação são diferentes, mas em geral, apresentam empate quando considerado o desvio-padrão. O teste de postos de Wilcoxon, representado no final da tabela por WS , sugere que CBF possui diferença estatística significativa se comparado as abordagens RSWV, QBWWV, BWWV e BC. O teste de Wilcoxon apresenta confiança de 95% (nível de significância de $\alpha = 0.05$) e a diferença, quando significativa, tem o sinal positivo (+) antecedendo o valor de p . A abordagem CBF apresenta melhoria na classificação se comparada a estes métodos, destacando-se as bases Faults, Glass, Heart, Laryngeal1, Laryngeal3, Liver, Sonar, Vehicle, Vertebral, WBC, WDVG, Weaning e Wine. Como estas bases apresentam duas ou mais classes e uma grande variabilidade na quantidade de atributos e na quantidade de observações, a abordagem proposta apresenta-se flexível quanto as possíveis aplicações.

A abordagem proposta *CBF: Grau de Entrada* N_{10} também foi comparada de modo pareado com as outras abordagens da literatura mencionadas e utilizando o teste de sinais sugerido em (DEMŠAR, 2006). Na Figura 15 está presente o número de vitórias, empates e derrotas agrupados por colunas que refletem a comparação da proposta com o método da literatura nomeado logo abaixo da respectiva coluna. A linha pontilhada refere-se ao teste de sinais, o qual a partir do total de problemas avaliados (n_{exp}). A hipótese nula H_0 sinaliza métodos estatisticamente equivalentes. Caso contrário, a rejeição da hipótese sugere que uma abordagem é melhor que a outra. A Equação (5.1) apresenta como obter o valor crítico (cv). Obtêm-se o valor $cv = 19.51$ a partir de $n_{exp} = 30$ com $\alpha = 0.05$ ($z_\alpha = 1.645$) (CRUZ; SABOURIN; CAVALCANTI, 2017). O método *CBF: Grau de Entrada* N_{10} apresenta resultados significativamente melhores que a literatura. De acordo com cv , todos os métodos da literatura, exceto por PW e PV, têm seus desempenhos quanto à taxa de reconhecimento abaixo de CBF, pois a quantidade de vitórias é maior que cv . Note que a quantidade de empates, por ser o mesmo desempenho dos dois métodos, exige que CBF tenha a linha de cv passando pelo número de vitórias para ser estatisticamente melhor que o método ao qual está sendo comparado. Em resumo, para o conjunto de experimentos realizados (30 problemas multiplicado por 9 métodos da literatura), CBF

ganhou em 198 de 270 casos (73.33%), perdeu em 57 casos (21.11%) e empatou em 15 casos (5.56%).

$$cv = \frac{n_{exp}}{2} + z_{\alpha} \times \frac{\sqrt{n_{exp}}}{2} \quad (5.1)$$

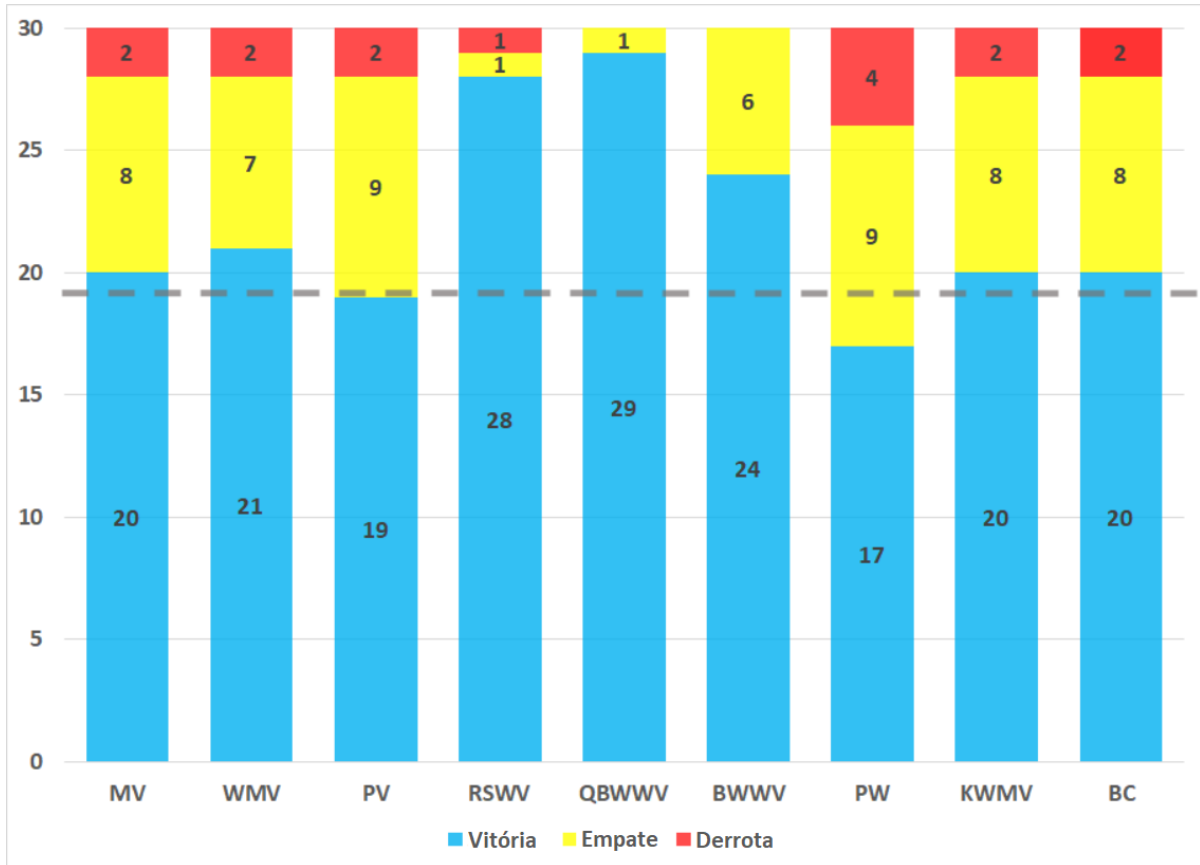


Figura 15 – Comparação pareada de CBF: Grau de Entrada N_{10} com métodos da literatura.

A Figura 16 apresenta uma análise estatística usando o teste de Friedman e o teste post hoc Nemenyi. Como pode ser visto, o método proposto é estatisticamente semelhante aos métodos da literatura PW, PV, KWMV, MV, WMV e BC, de acordo com a distância crítica (CD). Entretanto, o método CBF diferencia-se de BWWV, QBWWV e RSWV segundo o mesmo critério.

5.1.4 Discussão acerca de CBF

Esta seção procurou responder questões relacionadas ao uso de medidas de centralidade para estimar a importância dos classificadores como uma estratégia para a fusão estática de um conjunto de classificadores. Um primeiro questionamento levantado refere-se à capacidade das medidas de centralidade contribuírem para a acurácia de um

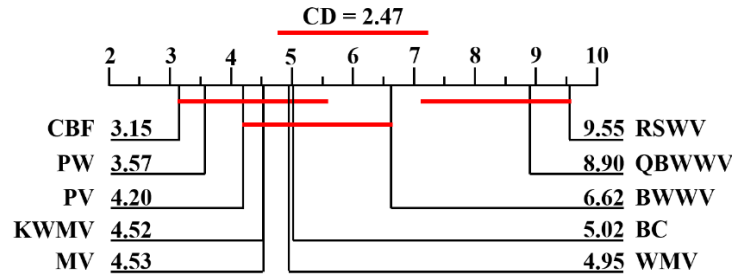


Figura 16 – Teste de Friedman e teste post-hoc Nemenyi comparando *CBF: Grau de Entrada N_{10}* com métodos da literatura.

conjunto de classificadores. Para responder a esse questionamento, houve a necessidade de representação do conjunto de classificadores como uma rede, a qual exigiu algum tipo de relacionamento pareado para representação de arestas. O método proposto, chamado de CBF, foi analisado quanto ao uso de diferentes medidas de centralidade combinadas a diferentes tipos de relacionamento, afinal, a centralidade é responsável por estimar a importância dos membros da rede pela análise dos vários relacionamentos pareados existentes entre os objetos. A melhor combinação entre relacionamento e centralidade, *CBF: Grau de Entrada N_{10}* , teve sua acurácia média comparada a literatura. Considerando 9 métodos diferentes (MV, WMV, PV, RSWV, QBWWV, BWWV, PW, KWMV e BC), a abordagem foi capaz de obter o melhor desempenho (maior acurácia média) em 73.33% dos casos enquanto perdeu apenas em 21.11%. O melhor concorrente, PW, foi melhor que *CBF: Indegree N_{10}* em apenas 4 de 30 problemas, perdendo em 17, e empatando em 9.

A literatura mostra que o relacionamento entre as medidas de diversidade e a acurácia do conjunto (KUNCHEVA; WHITAKER, 2003) possuem pouca correlação, por isso, estudos recentes sugerem a estimativa da boa e da má diversidade para o conjunto (BROWN; KUNCHEVA, 2010). Neste estudo, foram analisadas diferentes relações entre classificadores. Os relacionamentos avaliados foram *Falta Dupla*, *Estatística Q*, *Coefficiente de Correlação*, *Desacordo*, *Estatística Kappa* e N_{10} . Para analisar estes relacionamentos e obter qual deles é mais interessante para a acurácia do conjunto de classificadores, foram empregadas diversas medidas de centralidade, as quais pontuaram a importância dos classificadores pela análise de diversos relacionamentos, resultando em 21 cenários para CBF. A pontuação obtida pela centralidade foi utilizada como um método de ponderação do voto dos classificadores. Isto revelou que os relacionamentos mais interessantes foram o direcionado N_{10} e o não-direcionado *Falta Dupla*. O primeiro relacionamento, N_{10} , foi o único assimétrico avaliado, logo, a medida de centralidade deveria considerar esse aspecto, portanto, a *Centralidade de Grau de Entrada* foi escolhida. Essa combinação se mostrou

interessante visto que classificadores que mais complementaram as decisões de seus pares receberam maior peso quanto a seu voto. Além disso, a correlação de Pearson mostrou que o peso tem correlação positiva com a acurácia, mas não a ponto de serem considerados iguais. A medida de diversidade *Falta Dupla* constituiu o relacionamento simétrico mais importante. Portanto, reduzir o peso dos classificadores que mais erravam os mesmos objetos foi preferível a identificar somente a divergência entre opiniões como é o caso das medidas de diversidade *Desacordo*, *Estatística Q*, *Coeficiente de Correlação* e *Estatística Kappa*.

Ao avaliar a melhor medida de centralidade para o método de fusão CBF, considera-se também o relacionamento entre pares de classificadores, pois é a partir dos relacionamentos que a centralidade é obtida. O peso do classificador, obtido pela pontuação de centralidade, está associado com a influência de cada classificador no processo de rotulação de novos padrões, logo, a concorrência entre os classificadores é feita mediante analisando o relacionamento de cada classificador para todos os demais.

De modo geral, percebe-se que 3 medidas de centralidade possuem um posicionamento médio bastante similar. A *Centralidade de Autovetor* e a *Centralidade de Grau* se destacaram em relações simétricas enquanto que a *Centralidade de Grau de Entrada* foi a única medida avaliada para relações assimétricas, mas foi a que forneceu em média, os melhores resultados. A *Centralidade de Grau de Entrada* estima pelo número de relacionamentos N_{10} direcionados ao classificador correto, o quanto este complementa a decisão dos demais no conjunto. Um classificador com maior influência no voto neste caso é um classificador que frequentemente está certo em casos que seus pares não estão. A *Centralidade de Grau* é obtida pela proporção de casos onde um classificador em particular evita estar errado com seus pares, isto é, associa-se um peso maior para classificadores com baixo valor de N_{00} . Desta forma, esse classificador valoriza os outros 3 tipos de relacionamento que são N_{11} , N_{10} e N_{01} , ou seja, há pelo menos um dos dois classificadores correto.

Embora a medida de *Centralidade de Grau* seja atrativa, a medida de *Centralidade de Autovetor* para o relacionamento *Falta Dupla* teve um posicionamento um pouco melhor. Enquanto *Grau* apenas considerou os vizinhos diretos do classificador na rede, *Autovetor* considerou também os vizinhos diretos de seus vizinhos diretos. Portanto, o cenário da *Centralidade de Grau*, que é local, apresenta bons resultados. Entretanto, o cenário da *Centralidade de Autovetor*, mais abrangente, apresenta melhores resultados que a *Centralidade de Grau*.

Uma ressalva quanto ao tipo de rede de classificadores criadas precisa ser feita. Quando cada classificador da rede tem relacionamento com todos os demais, tem-se uma rede completa. Logo, a centralidade de *Grau* tida como uma medida local acaba avaliando uma quantidade grande de relacionamentos que dificilmente seriam avaliadas em outras redes que representam problemas do mundo real. Outro problema é a limitação da estimativa da centralidade de *Intermediação*, pois se a diferença entre os relacionamentos não é grande a ponto de envolver mais que 2 classificadores nos geodésicos, contar a quantidade de vezes que o classificador esteve presente entre os classificadores dos geodésicos é impossível. Uma alternativa para possibilitar a estimativa de centralidades baseadas em geodésico é a simplificação de rede.

Para que fosse possível propor um novo método de fusão, foi realizado um estudo de base que envolveu primeiramente diversas combinações entre o relacionamento representado na rede e a centralidade a qual iria apontar a importância do classificador pela análise dos relacionamentos dele na rede. Após esse estudo é que foi possível comparar o método com a literatura. Para responder a estes questionamentos também foi importante estabelecer um protocolo experimental que utilizasse diferentes problemas de classificação, para que pudesse ser avaliado o desempenho do método perante uma variedade de cenários. Portanto, 30 problemas de classificação foram avaliados, levando a criação de 180 conjuntos diferentes considerando a validação cruzada com 6-fold. O par de parâmetros relacionamento/centralidade, os quais não poderiam ser avaliados separadamente, gerou 21 situações, logo, 3780 conjuntos ponderados diferentemente foram avaliados ao todo.

Com um protocolo experimental robusto, CBF foi comparado a 9 métodos de fusão ponderada. Esta comparação mostrou que o método proposto é competitivo e, além disso, é distinto da maioria se considerada a acurácia média das abordagens é comparada de forma pareada. Os testes estatísticos mostraram que das 9 abordagens, CBF tem acurácia média significativamente superior a 3 métodos pelo teste post hoc de Nemenyi e 4 pelo teste de sinais de Wilcoxon. Na contagem dos melhores resultados estimados, CBF alcançou 14 entre os 30 problemas, enquanto que o melhor concorrente somente atingiu 6.

5.2 Avaliação experimental de CBS

O método nomeado como CBS é avaliado nesta seção. Esta abordagem é descrita como uma estratégia de seleção gulosa a qual é influenciada pela importância dos classificadores dispostos em rede. O protocolo experimental usa as mesmas divisões de

CBF, ou seja, a partir do conjunto original S tem-se os conjuntos de treinamento S_{train} , validação S_{val} e teste S_{test} . O conjunto de treinamento é usado para a criação do conjunto de classificadores. O conjunto de validação é usado para a avaliação do relacionamento pareado entre os classificadores. Com os relacionamentos já estimados, é possível criar a Rede de Classificadores e a partir dela, estimar a importância dos classificadores. Todo o processo de escolha foi obtido a partir da análise de um classificador candidato c_i em união com o subconjunto alvo C' . Os métodos da literatura seguiram o mesmo protocolo, ou seja, sempre que houve a necessidade de avaliar a acurácia ou diversidade do subconjunto objetivo C' , o subconjunto foi avaliado usando uma base de validação.

5.2.1 Comparação com Subconjunto Ótimo

Nesta seção foi utilizada a base de dados P2Problem (VALENTINI, 2005; ROY et al., 2018) com 500 instâncias. Esta base tem somente dois atributos numéricos e duas classes e tem sido usada para explicar o desempenho de soluções diversas de classificação devido seu complexo limite de decisão. Cada traçado no gráfico corresponde a uma função. A referida base pode ser observada na Figura 17.

A base P2Problem foi usada para criar um pool de $T = 20$ classificadores para relacionar a centralidade dos classificadores a um subconjunto ótimo OC' , determinado por uma estratégia de seleção por força bruta. Desta forma, todas as combinações possíveis foram feitas e em seguida a taxa de reconhecimento delas foi estimada em uma base de teste S_{test} . O objetivo foi identificar como a centralidade do classificador estimada na base de validação S_{val} poderia estar relacionada com o cenário ideal de teste. Para isso, após estimar a centralidade dos classificadores, eles foram ordenados de forma ascendente como é possível notar na Tabela 9.

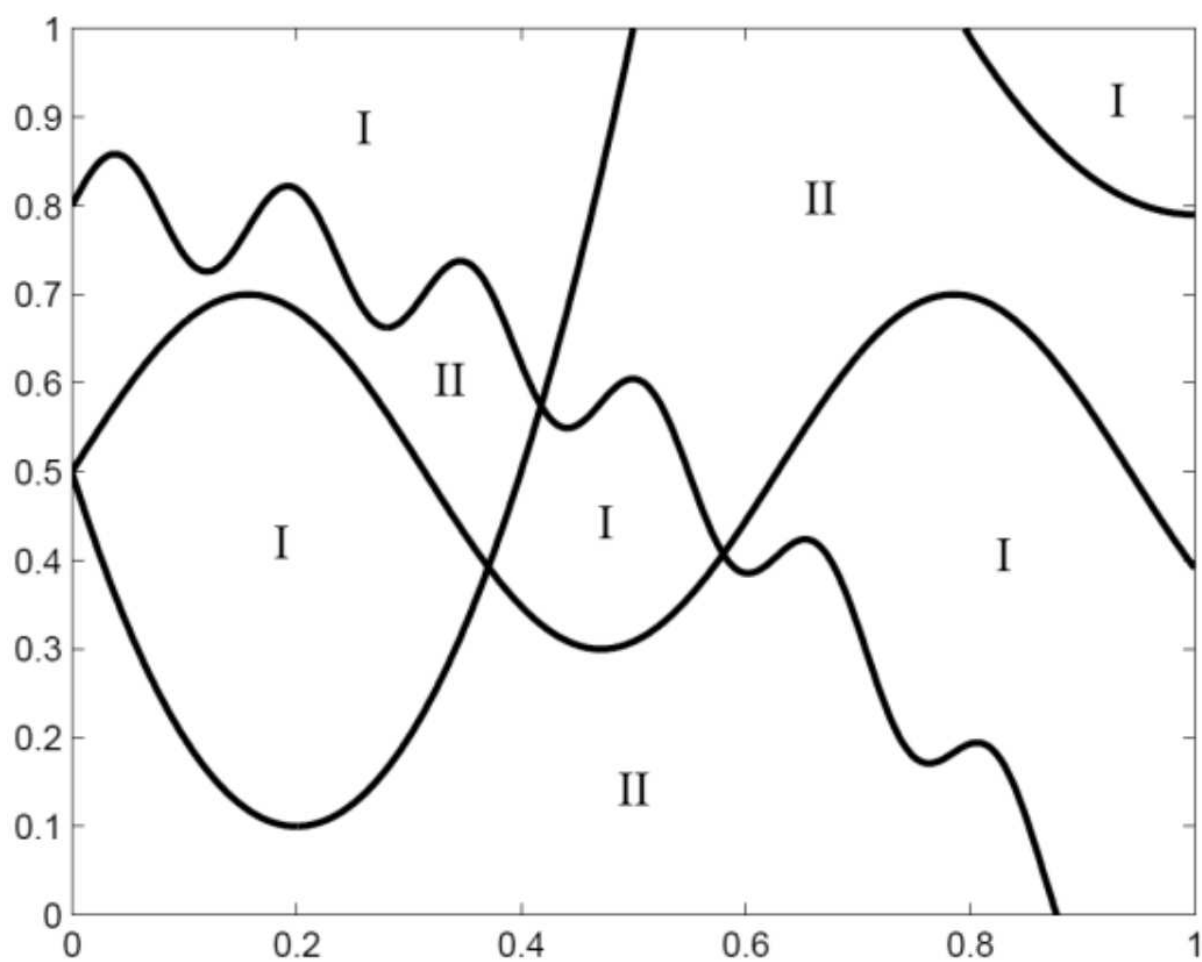


Figura 17 – Distribuição das classes (I e II) para a base de dados P2Problem.

Tabela 9 – Classificadores gerados por Bagging ordenados pelo peso da *Centralidade de Grau de Entrada* em uma rede de relacionamentos N_{10} . (*) significa que o classificador pertence ao subgrupo ótimo.

id	valor
16	188
1	191
7	201
18	206
0	208
2*	219
10	222
5	260
19	267
17	283
12	285
11	304
15	304
9	307
4	325
6	337
13	337
3*	348
14	358
8	369

Tabela 10 – Classificadores gerados por Boosting com validação cruzada (k-fold com $k = 6$) ordenados pelo peso da *Centralidade de Grau de Entrada* em uma rede de relacionamentos N_{10} . (*) significa que o classificador pertence ao subgrupo ótimo.

fold 0		fold 1		fold 2		fold 3		fold 4		fold 5	
id	valor	id	valor	id	valor	id	valor	id	valor	id	valor
18	308	14	336	3	291	6	262	4	460	12	249
11	349	0	339	16	325	13	329	5	461	2	261
0	355	2	339	14	354	3	336	14	482	10	273
6	355	5	339	17	392	8	336	17	510	7	279
8	355	6	339	5	407	9	336	2	516	11	279
10	355	8	339	7	407	18	336	7	523	13	279
12	355	10	339	9	407	12	380	10	523	17	279
17	355	12	339	15	407	4	422	12	523	18	279
19	355	7	361	19	407	17	423	6	528	8	282
15	459	4*	409	1	420	14	446	16	540	16	292
16	481	13	451	0*	482	11	478	13	676	15	380
2	521	19	461	8	503	5	492	11	680	14	401
14*	550	16*	477	2	586	10*	559	9	702	0*	491
4*	568	11	506	18	603	7	599	0*	707	6	523
9*	674	1	528	13	622	16	606	15*	710	5	530
7*	773	15	604	11	659	0*	661	19	758	4	532
3	807	17	618	12	688	2	675	3	780	19*	603
13*	826	3*	807	10	752	19*	759	8	844	1*	730
1*	841	9	817	4*	835	1*	816	18*	1038	9*	907
5	928	18*	1188	6*	1022	15	885	1	1051	3	926

A Tabela 9 mostra a identificação do classificador (id) e o peso associado a *Centralidade de Grau de Entrada* numa rede cujo relacionamento é N_{10} . Os classificadores 2 e 3 compõe o subconjunto ótimo, logo, $OC' = \{c_2, c_3\}$. O caso mostrado na Tabela 9 se repetiu entre os 6 folds analisados e por consequência não houve um padrão entre o valor da centralidade do classificador e a participação dele em OC' . Entretanto, para o mesmo problema (P2Problem) quando o mesmo processo foi realizado numa base de classificadores gerados por *Boosting* revelou-se um padrão interessante. A Tabela 10 apresenta os classificadores gerados por Boosting, cujo relacionamento pareado é dado por N_{10} e a análise de centralidade é pelo *Grau de Entrada*. Os classificadores marcados com asterisco são os classificadores que formam o subconjunto ótimo. Há uma tendência desse subconjunto ótimo ser formado pelos 50% classificadores mais importantes. Esse mesmo padrão foi observado em outros problemas para diferentes medidas de centralidade (exceto pela *Centralidade de Intermediação*) para todos os outros relacionamentos pareados avaliados (*Estatística Q*, *Coeficiente de Correlação*, *Desacordo*, *Falta Dupla* e *Estatística Kappa*). Para lembrar, o problema da Centralidade de Intermediação na estimativa de importância é que precisa haver pelo menos um classificador entre o classificador inicial e o classificador final usados para calcular o geodésico. Se a diferença entre o relacionamento pareado numa rede completa não envolver pelo menos 3 classificadores nos geodésicos, a Centralidade de Intermediação gera valor nulo.

Houve um padrão de posicionamento dos classificadores pertencentes a OC' para a geração de classificadores feitas pelo método Boosting na base P2Problem. Ao ordenar a *centralidade de Grau de Entrada*, os classificadores de OC' frequentemente estão presentes entre os classificadores de maior centralidade. O índice do classificador pertencente a OC' e, menos central, geralmente corresponde a um índice inferior a 50% do tamanho total do conjunto. Isto sugere que é possível extrair o subconjunto ótimo a partir da análise de apenas parte do conjunto inicial. Portanto, o espaço de busca relacionado ao parâmetro p do algoritmo proposto *CBS* pode ser de 100%, mas há a possibilidade de encontrar OC' no espaço reduzido de 50%. O desafio é então, na maior parte dos problemas, encontrar ao menos parte dos classificadores pertencentes a OC' e complementar o subconjunto escolhido com classificadores que poderiam formar soluções sub-ótimas melhores que as soluções encontradas na literatura.

Para encerrar esta seção, é importante reforçar que o método de Geração de Classificadores, seja Bagging ou Boosting, não é determinante para *CBS* encontrar o subconjunto de classificadores ótimo. Entretanto, como os demais métodos da literatura, o

desempenho de CBS varia de problema a problema.

5.2.2 Comparação de CBS com a literatura

Neste experimento foram utilizados a centralidade de Grau de Entrada para a análise da importância dos classificadores em uma rede cujo relacionamento entre os classificadores foi representada por N_{10} . Todos os outros parâmetros referem-se a natureza da criação do pool, como o tamanho do conjunto $T = 100$ e o classificador base escolhido foi o Perceptron de Erro Quadrado Mínimo, o mesmo usado para avaliar a proposta de fusão. O algoritmo de geração de classificadores, entretanto, é o Boosting. A grande vantagem apresentada por esta estratégia é a indução da complementariedade do conjunto. A complementariedade, por sua vez, é uma das medidas mais utilizadas em métodos de seleção gananciosa (AHMED et al., 2018; PARTALAS; TSOUMAKAS; VLAHAVAS, 2012). Embora o *Boosting* seja sugerido em problemas de seleção gananciosa, CBS e a literatura relacionada apresentam flexibilidade neste quesito e, portanto, não estão restritas a esse método de geração de classificadores em particular. Entretanto, o método de geração de classificadores pode influenciar o processo de seleção gulosa, como é apresentado no Apêndice A.

Os métodos da literatura utilizados para comparação foram o AGOB (MARTÍNEZ-MUÑOZ et al., 2004), DREP (LI; YU; ZHOU, 2012), Kappa Pruning (MARGINEANTU; DIETTERICH, 1997) e POBE (MARTÍNEZ-MUÑOZ; SUÁREZ, 2006). Além destes, o melhor classificador (*Single Best* - SB) e a combinação por voto majoritário (MV) foram avaliadas, servindo de base para medir as alternativas de seleção mais complexas. Enquanto o SB mostra a acurácia somente do melhor classificador (escolhido na fase de treinamento), MV mostra a acurácia do SMC sem a seleção de um subconjunto, logo, todos os membros exercem voto no processo de rotulação. Todos os métodos estão descritos na seção 3.2. A Tabela 11 apresenta a acurácia média dos métodos para os respectivos problemas de classificação e a média de classificadores usados. Em Kappa foi fixado o tamanho do subconjunto em metade do pool inicial e em DREP foi usado o parâmetro $p = 0.5$, o mesmo usado por CBS. A escolha de p foi devido as observações descritas na Seção 5.2.1.

Tabela 11 – Comparação de CBS com métodos da literatura. Os valores apresentados são a taxa de reconhecimento obtida por k-fold com $k = 6$. O valor ao lado ($|C'|$) refere-se ao tamanho médio do subconjunto selecionado. Em negrito está o melhor resultado daquele problema de classificação. # Best é quantidade de melhores resultados atingidos por um método. WS significa teste de sinais de Wilcoxon, sendo que o valor representado é referente ao valor de p, enquanto que (+) é para sinalizar resultados significativos com $\alpha = 0.05$.

	Kappa	C'	DREP	C'	POBE	C'	AGOB	C'	SB	C'	MV	C'	CBS	C'
Australian	84.50 ± 2.38	50	84.35 ± 2.66	4	83.90 ± 2.52	13	83.90 ± 2.55	5	83.19 ± 3.16	1	85.51 ± 2.59	100	85.07 ± 3.31	7
Banana	81.25 ± 2.24	50	82.95 ± 2.18	2	81.25 ± 2.21	26	80.25 ± 2.20	15	84.95 ± 2.23	1	84.05 ± 2.37	100	89.65 ± 2.64	9
Blood	78.22 ± 3.27	50	76.21 ± 3.41	7	78.22 ± 3.34	23	77.02 ± 3.35	22	78.34 ± 2.02	1	77.01 ± 4.46	100	78.61 ± 2.19	21
CTG	89.98 ± 1.41	50	89.09 ± 1.58	6	89.98 ± 1.49	16	91.38 ± 1.51	5	88.20 ± 1.00	1	89.79 ± 1.10	100	89.46 ± 1.51	8
Diabetes	73.89 ± 2.78	50	75.98 ± 3.13	3	73.29 ± 2.95	29	74.69 ± 2.99	17	76.77 ± 4.15	1	73.63 ± 2.92	100	76.51 ± 3.43	4
Ecoli	86.90 ± 4.33	50	86.31 ± 5.61	3	86.90 ± 4.97	25	88.20 ± 5.10	27	85.12 ± 5.52	1	87.20 ± 4.99	100	85.12 ± 4.80	9
Faults	68.73 ± 1.97	50	66.56 ± 1.90	2	60.53 ± 1.93	14	66.33 ± 1.92	2	68.31 ± 1.63	1	69.19 ± 1.66	100	68.98 ± 1.81	5
German	74.50 ± 1.75	50	71.70 ± 3.51	5	72.70 ± 2.63	8	69.80 ± 2.80	4	70.30 ± 3.32	1	73.90 ± 2.92	100	71.60 ± 2.11	6
Glass	64.03 ± 5.82	50	66.84 ± 7.43	5	64.63 ± 6.63	9	63.23 ± 6.79	10	58.92 ± 5.91	1	64.50 ± 7.66	100	65.88 ± 4.36	10
Haberman	61.44 ± 6.47	50	67.00 ± 5.12	3	60.44 ± 5.79	45	58.94 ± 5.66	41	72.55 ± 6.30	1	63.40 ± 8.22	100	71.90 ± 5.39	14
Heart	80.74 ± 4.38	50	78.15 ± 8.55	4	80.94 ± 6.47	14	80.84 ± 6.88	20	78.52 ± 10.24	1	83.33 ± 5.41	100	77.41 ± 8.26	10
ILPD	62.96 ± 4.53	50	62.26 ± 3.47	3	62.06 ± 4.00	23	62.46 ± 3.90	19	65.88 ± 3.82	1	60.73 ± 5.53	100	69.82 ± 2.44	7
Ionosphere	82.02 ± 4.62	50	79.46 ± 5.41	2	82.82 ± 5.02	32	79.62 ± 5.10	29	82.31 ± 4.01	1	82.30 ± 4.73	100	81.75 ± 3.04	8
Laryngeal1	80.77 ± 4.62	50	80.32 ± 5.30	5	78.47 ± 4.96	26	78.17 ± 5.03	23	80.36 ± 6.49	1	81.23 ± 4.40	100	78.92 ± 4.58	8
Laryngeal3	65.74 ± 6.89	50	68.57 ± 6.91	4	64.54 ± 6.90	10	65.14 ± 6.90	7	71.42 ± 6.66	1	69.41 ± 5.93	100	72.83 ± 6.43	8
Lithuanian	85.15 ± 1.39	50	85.45 ± 1.96	3	83.95 ± 1.67	5	85.75 ± 1.73	9	83.25 ± 1.75	1	83.35 ± 1.70	100	85.40 ± 1.97	3
Liver	63.50 ± 3.22	50	65.84 ± 6.22	4	63.30 ± 4.72	31	62.80 ± 5.02	26	69.02 ± 4.75	1	64.66 ± 3.43	100	68.14 ± 5.61	9
Magic	67.85 ± 0.25	50	70.67 ± 1.98	2	67.75 ± 1.12	26	68.65 ± 1.29	2	79.02 ± 0.85	1	67.46 ± 0.23	100	76.23 ± 0.67	2
Mammo	66.38 ± 3.97	50	77.83 ± 3.28	2	65.98 ± 3.62	23	66.18 ± 3.55	28	82.90 ± 2.83	1	78.31 ± 2.80	100	82.42 ± 2.69	6
Monk	78.24 ± 3.82	50	87.04 ± 2.36	4	80.54 ± 3.09	17	75.44 ± 2.94	4	86.81 ± 3.18	1	76.39 ± 4.25	100	92.13 ± 6.09	7
Phoneme	71.82 ± 2.50	50	76.24 ± 2.18	4	64.22 ± 2.34	23	74.82 ± 2.31	28	77.32 ± 1.52	1	75.04 ± 2.57	100	77.76 ± 1.40	7
Segmentation	95.07 ± 0.80	50	95.50 ± 0.80	6	94.01 ± 0.80	38	96.47 ± 0.80	35	94.55 ± 1.18	1	95.24 ± 1.10	100	95.50 ± 0.57	6
Sonar	75.06 ± 6.17	50	76.02 ± 5.45	4	77.47 ± 5.81	26	74.26 ± 5.74	22	72.13 ± 3.02	1	74.10 ± 7.04	100	75.52 ± 2.93	9
Thyroid	94.79 ± 1.89	50	92.62 ± 2.97	4	92.49 ± 2.43	35	94.29 ± 2.54	13	94.07 ± 2.28	1	95.08 ± 1.57	100	94.65 ± 1.98	8
Vehicle	80.02 ± 1.80	50	73.52 ± 3.63	3	78.82 ± 2.72	9	79.32 ± 2.90	5	77.19 ± 2.60	1	80.14 ± 2.42	100	75.30 ± 3.60	6
Vertebral	86.00 ± 3.83	50	84.00 ± 2.00	11	85.90 ± 2.91	12	84.80 ± 2.73	13	83.67 ± 3.14	1	86.33 ± 3.54	100	84.67 ± 2.75	17
WBC	97.19 ± 1.68	50	96.49 ± 0.99	7	97.39 ± 1.34	11	96.59 ± 1.27	8	96.49 ± 0.99	1	97.01 ± 1.41	100	96.84 ± 1.21	21
WDVG	83.72 ± 1.14	50	84.38 ± 0.75	3	85.01 ± 0.95	5	84.52 ± 0.91	4	85.36 ± 1.68	1	83.30 ± 1.08	100	85.66 ± 1.70	4
Weaning	77.46 ± 4.79	50	74.54 ± 5.90	2	77.66 ± 5.34	28	77.56 ± 5.46	24	79.80 ± 6.82	1	79.45 ± 4.34	100	79.47 ± 3.76	7
Wine	97.19 ± 1.26	50	98.32 ± 1.69	23	97.19 ± 1.47	23	97.19 ± 1.52	17	96.65 ± 1.93	1	96.61 ± 1.99	100	96.61 ± 1.99	65
# Best	1		2		3		4		6		7		7	
WS	+ 0.03318		+ 0.0012		+ 0.00932		+ 0.00452		+ 0.0455		+ 0.0601		N/A	

A Tabela 11 apresenta os melhores resultados em negrito (# Best). Percebe-se que o método CBS tem destaque, pois tem os melhores resultados em 7 de 30 problemas analisados, empatando apenas com o MV, que não seleciona classificadores. Destes 7 problemas em que CBS se destaca, apenas uma base de dados, a Ecoli, possui mais de 2 classes. O terceiro melhor concorrente foi o Single Best (SB) com 6 melhores resultados. Ao final da tabela encontra-se o valor de p , obtido pelo teste de sinais de Wilcoxon. Portanto, CBS apresentou diferença estatística suficiente em relação todos os demais como pode ser visto pelo sinal (+). A confiança deste teste é de 95% e o valor de p é significativo se $p \leq 0.05$.

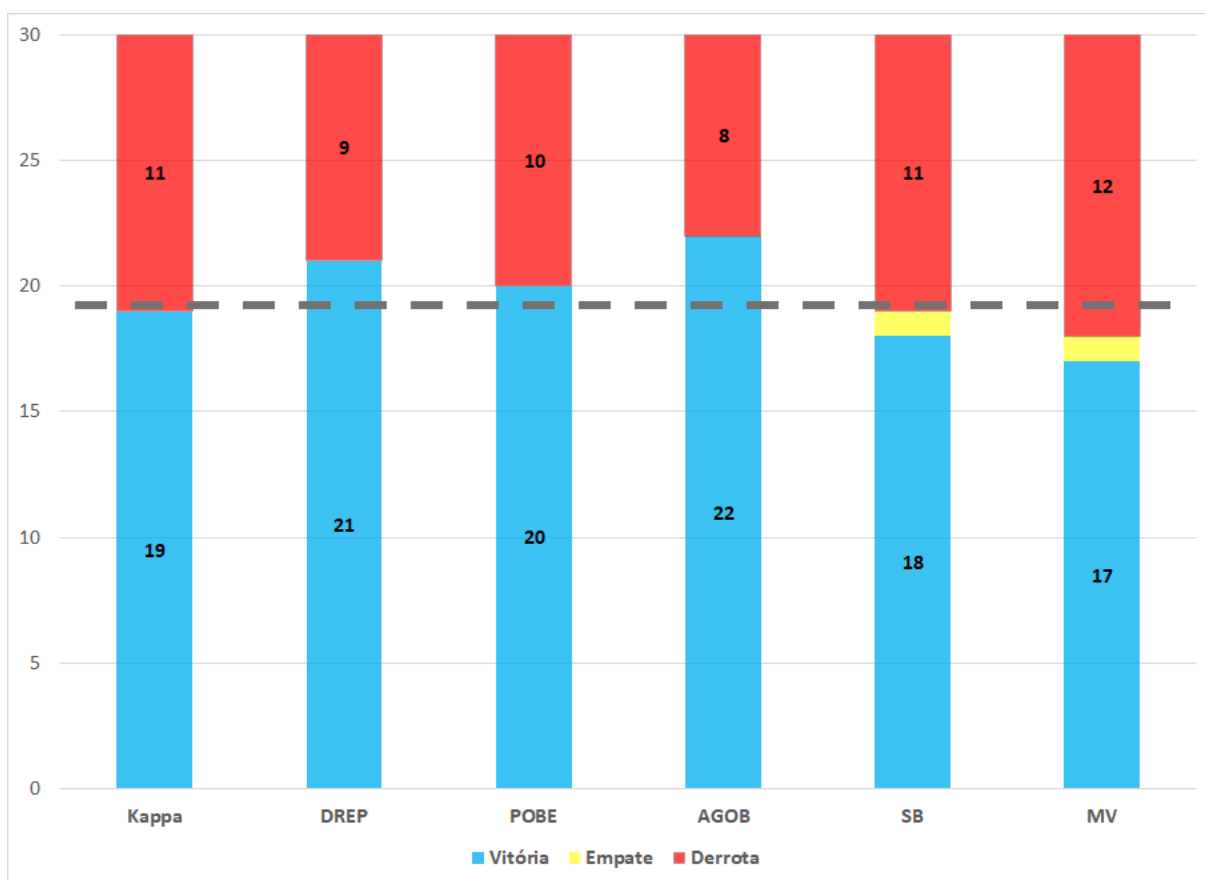


Figura 18 – Comparação pareada de CBS com a literatura.

A base Wine diferenciou-se bastante dos demais problemas quanto ao número de classificadores selecionados em média na abordagem CBF. Para entender o motivo, foi avaliado o peso dos relacionamentos da rede. O peso da rede direcionada N_{10} aponta a quantidade de instâncias que um classificador acerta enquanto seu par erra. Para os 9900 relacionamentos entre os 100 classificadores, observou-se que em média a quantidade de instâncias que um classificador complementa seus pares é de apenas 1.16 com desvio-padrão de 1.06. Portanto, esses classificadores gerados para a base Wine pouco contribuem entre

si, gerando uma falsa expectativa em CBS que ao adicionar mais classificadores aumente a taxa de reconhecimento do subconjunto.

Uma outra análise foi feita comparando estatisticamente a abordagem nova CBS com a literatura. A Figura 19 apresenta a comparação da posição média de cada solução sendo que quanto mais baixo melhor, ou seja, mais vezes o método esteve nas primeiras posições. No caso da abordagem proposta, ela costuma estar entre as primeiras posições e diferencia-se das abordagens POBE e AGOB.

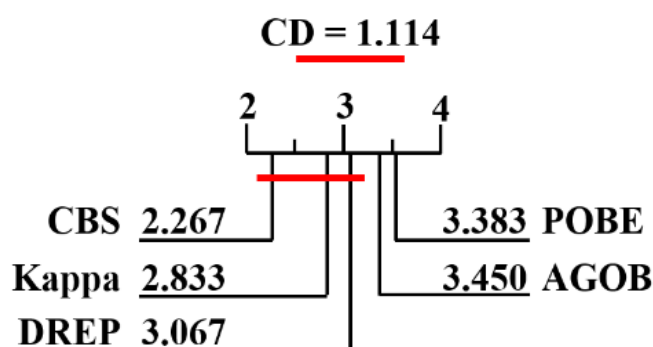


Figura 19 – Teste de Friedman e post hoc Nemenyi comparando CBS com a Literatura.

A Figura 18 apresenta uma comparação entre o método CBS e o método identificado na coluna. O traço é o valor crítico (cv) estimado da mesma maneira como apresentado nos experimentos com fusão de classificadores. Portanto, para 30 problemas de classificação, $cv = 19.5$. Assim, o método apresentou-se suficientemente diferente da literatura, exceto por Kappa Pruning, pois para ser diferente dos métodos precisava de pelo menos 20 vitórias (cv com arredondamento).

A Figura 20 apresenta a acurácia média de cada abordagem representada no eixo X e também a quantidade de classificadores (em média) representado no eixo Y. Percebe-se que a nova proposta, CBS, além de ter a taxa de reconhecimento mais precisa, possui uma quantidade bastante reduzida de classificadores compondo o subconjunto. Quanto ao último quesito, somente perde para o DREP, portanto, os classificadores que a abordagem CBS apresenta a mais em relação ao DREP ajudam o ensemble selecionado estaticamente a ser mais preciso.

5.2.3 Discussão acerca de CBS

O CBS mostrou-se uma alternativa atraente frente a literatura. O maior atrativo deste algoritmo é a redução do espaço de busca pelo subconjunto ótimo quando o método de criação do pool de classificadores é o *Boosting*. A quantidade de classificadores selecionados a

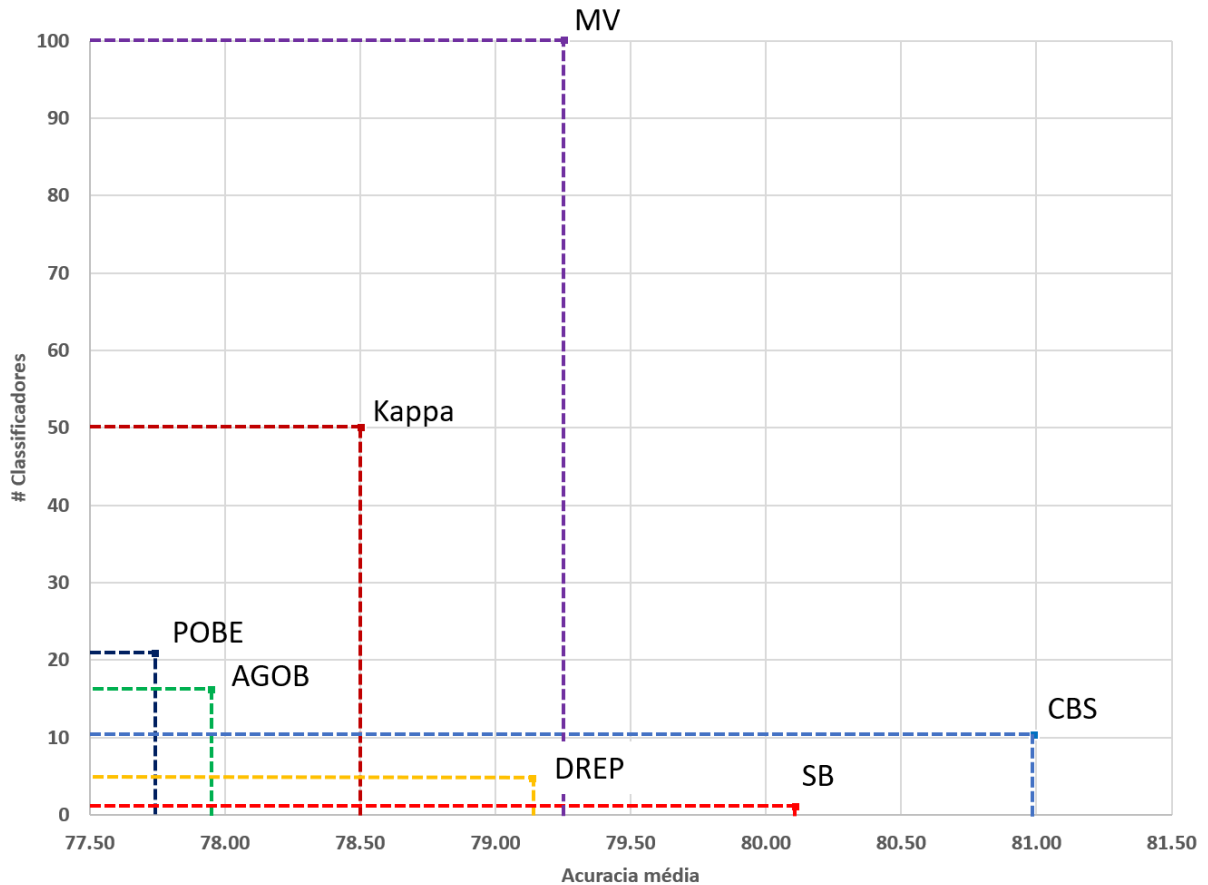


Figura 20 – Comparação entre abordagens de seleção quanto à acurácia média e o tamanho médio do conjunto de classificadores gerados por Boosting.

partir do pool inicial geralmente é inferior a 20% do conjunto inicial. Para chegar no melhor subconjunto, CBF ordena os classificadores pela capacidade que eles tem de complementar seus pares. Foi encontrado empiricamente que cerca de 50% dos classificadores do conjunto estão entre os mais importantes e a partir destes, até 20 poderão compor o conjunto final. A única exceção foi para a base Wine, com 65 classificadores. Isto aconteceu porque o parâmetro p sempre escolhe 50% dos classificadores do conjunto restante e não apenas 50 classificadores do conjunto inicial. Como o critério de parada de CBS é pela perda de acurácia do subconjunto C' , o método adiciona classificadores na esperança de que pequenas diferenças possam ser encontradas pela maioria dos classificadores. Esse problema de classificação especificamente (Wine) é um dos mais difíceis estudados, pois os classificadores criados diferenciavam-se muito pouco uns dos outros. A média dos relacionamentos N_{10} foi cerca de 1.34 com desvio-padrão de 0.95. Isto quer dizer que os relacionamentos entre pares de classificadores apenas eram capazes de complementar as decisões de seus pares em apenas uma a duas instâncias. A comparação indica que CBS superou a literatura em 82 casos (68.33%) e perdeu em 38 casos (31.67%).

Os métodos de seleção gananciosa diferenciam-se em alguns aspectos básicos como a ordem a qual o classificador é avaliado antes de ser adicionado ao subconjunto, a medida pela qual os classificadores são ordenados, o critério de seleção e se quantidade de classificadores do subconjunto é fixa ou flexível. Primeiramente, CBS ordenou os classificadores de acordo com o importância que os classificadores tem em complementar a falha de seus pares. Os demais métodos da literatura também usaram o critério da complementariedade, cada um a seu modo, pois para alguns a complementariedade de um classificador era em relação ao subconjunto e para outros era em relação a seus pares individualmente. No critério de seleção, CBS escolhe pela acurácia do candidato, semelhantemente ao que DREP faz. Inclusive, ambas as abordagens avaliam a acurácia apenas dos N primeiros classificadores da lista ordenada - segundo o critério particular da complementariedade do candidato para o subconjunto. Kappa seleciona apenas pela ordem dos classificadores não empregando nenhum teste adicional.

A quantidade de classificadores a serem selecionados é fixa somente em Kappa Pruning, portanto, para DREP, AGOB, POBE e CBS a quantidade de classificadores que compõe o subconjunto é determinada dinamicamente na fase de treinamento. Desta forma, ocorre a seleção estática dos classificadores, pois não há mudança no subconjunto na fase de teste. O critério adotado em Kappa de selecionar 50% dos classificadores do conjunto foi obtido a partir da observação de que o subconjunto ótimo geralmente está inserido entre os 50% classificadores mais importantes do conjunto gerado, discutido na Seção 5.2.1. Da mesma forma foi limitado o espaço de busca para DREP e CBF a 50% dos possíveis candidatos (que ainda não se encontravam inseridos no subconjunto C'). Como p nas abordagens DREP e CBS é um valor arbitrário, a associação com o subconjunto ótimo obtido pela exaustiva observação de todas as combinações possíveis no contexto do problema de combinação P2Problem indica um caminho.

É importante mencionar que o relacionamento entre a importância dos classificadores gerados por Bagging e Boosting com o subconjunto ótimo diferem. O processo aleatório de composição dos classificadores obtidos por Bagging não indicou um relacionamento claro com o subconjunto ótimo, enquanto que para o Boosting houve. Mesmo assim, os resultados obtidos para o experimento com Bagging sugere que a abordagem CBS é atraente se comparada a literatura. A importância do classificador quanto a capacidade de complementar a decisão de seus pares, entretanto, não pode ser o único critério de seleção. Por isso, tanto DREP quanto CBS avaliam dentre os classificadores que mais complementam o subconjunto, aqueles que quando inseridos neste subconjunto poderiam

produzir os melhores resultados e assim decidem pela inclusão somente do melhor (mais acurado) a cada iteração.

Este trabalho apresentou um novo critério para a ponderação dos classificadores que pode ser usado tanto para a fusão quanto para a seleção. O critério da complementariedade estimada a partir de uma rede de classificadores se mostrou um avanço interessante. Além disso, a análise de centralidade dos classificadores preenche algumas lacunas deixadas por (TRAWINSKI; CORDON, 2016), que sugere que seria importante o estudo de diferentes centralidades estimadas em representações de relacionamento variadas para o mesmo conjunto associando com a acurácia. O autor sugere a utilização deste critério (centralidade) para a seleção de grupo ótimo de classificadores, portanto, o método CBS deu um passo para o avanço da área nestes quesitos. Foi apresentando pela abordagem CBS um subconjunto de classificadores cujo resultado sub-ótimo foi obtido pela busca em um espaço reduzido. Este espaço foi delimitado pelo relacionamento entre os classificadores do conjunto ótimo e os classificadores mais centrais na rede. Para chegar no conjunto sub-ótimo foram realizados experimentos diversos envolvendo a combinação de diversas medidas de centralidades, estimadas em diferentes tipos de relacionamento pareados. Do ponto de vista prático, os classificadores quando ordenados pelas respectivas centralidades raramente apresentaram sequências iguais. Entretanto, os classificadores mais importantes para uma medida raramente não foram importantes para a outra.

6 Conclusão

Este trabalho trouxe novas abordagens para a seleção e fusão de classificadores, baseando-se em medidas de centralidade no contexto das Redes Complexas. Ambas as propostas surgem a partir da análise de uma rede de classificadores, na qual vértices representaram os classificadores e arestas representaram o relacionamento entre os possíveis pares de classificadores. A avaliação dos classificadores neste contexto contribui para o objetivo geral do trabalho e, por consequência, para o avanço da área de Reconhecimento de Padrões.

O desafio de avaliar diferentes relacionamentos para cada rede de classificadores e o sucesso na ponderação sugerida pelas medidas de centralidade implicaram na criação de um protocolo experimental robusto. Neste protocolo, 30 problemas de classificação foram utilizados para confirmar nossa principal hipótese: a centralidade é uma estratégia de ponderação de classificadores interessante tanto para a combinação quanto para a seleção de classificadores. Para isso, primeiramente, a análise de diversos tipos de relacionamentos na rede foram realizadas por diferentes medidas de centralidade no contexto da fusão estática de classificadores. Foi observado que o relacionamento que mais esteve relacionado com a acurácia do conjunto foi o N_{10} , seguido do relacionamento *Falta Dupla*. As medidas de centralidade que se destacaram foram a *Autovetor* e o *Grau*. A comparação com a literatura foi feita pelo melhor par de parâmetros, isto é, o relacionamento N_{10} analisado pela centralidade de *Grau de Entrada*. Dos 30 problemas analisados, este método apresentou o melhor resultado em 14, permanecendo competitivo nos demais enquanto que o melhor concorrente apresentou apenas 6 melhores resultados. Os 9 métodos da literatura foram utilizados para realizar 270 comparações pareadas com esta proposta. Em 198(73.33%) casos o método ganhou, empatou em 15(5.56%) e perdeu em 57 casos (21.11%).

Como há várias medidas de diversidade na literatura (KAPP; SABOURIN; MAUPIN, 2007; KUNCHEVA; WHITAKER, 2001; JOHANSSON; LÖFSTRÖM; BOSTRÖM, 2013), além de outras medidas de centralidade (LÜ et al., 2016; SOSNOWSKA; SKIBSKI, 2017), futuramente novas combinações destes parâmetros podem ser avaliados para o aumento na precisão do conjunto de classificadores no processo de rotulação. Em especial, a idéia de combinar diferentes medidas de diversidade proposta em (CAVALCANTI et al., 2016) pode revelar também diferentes observações úteis para a composição de CBF. Neste

trabalho, foram escolhidas medidas de centralidade clássicas devido elas terem inspirado uma gama de outras medidas, além de diferenciarem-se claramente entre si no aspecto teórico. A análise de uma rede de classificadores proposta neste trabalho complementa requisitos cuja literatura recente tem demandado (TRAWINSKI; CORDON, 2016) quanto a aplicabilidade no contexto do Conjunto de Classificadores, Aprendizado de Máquina e Reconhecimento de Padrões. Finalmente, entre as contribuições de CBF, destaca-se a análise da rede assimétrica N_{10} pela *Centralidade de Grau de Entrada* como um método alternativo bastante atrativo a ponderação de acurácia.

Quanto ao método de seleção proposto, o CBS, foi explorada prioritariamente a capacidade de complementação de um classificador, obtido pela análise da *Centralidade de Grau de Entrada* numa rede de classificadores cujos relacionamentos foram representados por N_{10} . A escolha pela complementariedade se deu inspirada na literatura de seleção gulosa/gananciosa, porém, trazendo uma abordagem para o peso do classificador inspirada na importância que o classificador têm para a rede. O uso de rede para métodos de seleção estática ainda é uma idéia pouco explorada segundo (TRAWINSKI; CORDON, 2016).

Uma observação interessante houve na comparação de Bagging com Boosting quanto ao peso dos classificadores no processo de seleção estática: a centralidade do classificador é capaz de reduzir em cerca de 50% o espaço de busca do problema de classificação P2Problem. Diante deste cenário, o método proposto e os métodos da literatura foram ajustados a considerarem em seu processo de escolha somente metade do pool inicial de classificadores, ordenados pela métrica partilamente adotada por cada abordagem. CBS apresentou um resultado atrativo comparado a literatura e foi capaz de trazer os melhores resultados para 15 dos 30 problemas de classificação, enquanto que seu melhor concorrente neste critério apresentou somente 5. A abordagem nova de seleção superou a literatura em 82 casos (68.33%) e perdeu em 38 casos (31.67%). A quantidade de classificadores a comporem o subconjunto formado na fase de treinamento apresentou um custo interessante, em média, de 10 classificadores entre os 100 do conjunto original. Neste quesito, apenas DREP apresentou menos classificadores, entretanto, CBS apresentou a melhor taxa de reconhecimento. Isso sugere que a quantidade de classificadores influencia positivamente a taxa de reconhecimento.

O teste Friedman e o teste post hoc de Nemenyi indicam que CBS difere de AGOB e POBE enquanto é semelhante a Kappa Pruning e DREP. Entretanto, o teste de sinais de Wilcoxon indica que o método proposto de seleção difere significativamente dos demais avaliados. A comparação entre CBS e cada abordagem da literatura de forma pareada

revelou que o método proposto é significativamente diferente de DREP, POBE e AGOB, enquanto assemelha-se a Kappa. A comparação pareada levou em consideração a média dos classificadores para cada um dos 30 problemas avaliados e a diferença foi considerada significativa quando o número de vitórias do método foi acima do valor crítico.

Em média, ambos os métodos baseados em centralidade de rede - CBF para fusão e CBS para seleção - são superiores a literatura em aproximadamente 70% dos casos. Portanto, a análise de importância do classificador pela posição que ele ocupa na rede é uma alternativa promissora. Na próxima seção são apresentadas alternativas futuras para a evolução deste estudo.

6.1 Trabalhos Futuros

Este trabalho propôs que um Sistema de Múltiplos Classificadores fosse representado por uma rede e em seguida a análise de sua importância fosse estimada. Uma das contribuições foi as diretrizes para a representação da rede, a qual permitiu a análise correta da importância dos classificadores. A partir da importância dos classificadores, um método de fusão e outro de seleção foram propostos. Neste sentido, como sugestão de trabalho futuro está a aplicação da análise de rede de classificadores para a fase de geração de classificadores.

A rede de classificadores que foi representada e analisada estaticamente apresentou bons resultados para a fusão estática e a seleção estática de classificadores. A continuação desta pesquisa poderia encontrar no ambiente dinâmico de fusão e/ou seleção melhores resultados quanto a taxa de reconhecimento, pois para cada padrão avaliado poderia-se extrair pesos dinâmicos ou subconjuntos dinâmicos que reagissem ao padrão não rotulado e, por consequência, pudessem produzir uma resposta aperfeiçoada.

No protocolo experimental foram utilizadas bases de tamanho reduzido em comparação com bases de *Big Data*. Há de se investigar o desempenho dos métodos propostos para bases de dados maiores tanto na acurácia quanto no tempo de execução. Ainda sobre as bases de dados, deve ser investigadas as bases com presença de ruído.

Referências

AHMED, M. A. O. et al. Using Diversity for Classifier Ensemble Pruning: An Empirical Investigation. *Theoretical and Applied Informatics*, v. 29, n. 1, p. 25–39, 2018. ISSN 18965334.

ALCALÁ-FDEZ, J. et al. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, v. 17, n. 2-3, p. 255–287, 2011. ISSN 15423980. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-79951829331&partnerID=tZOtx3y1>>.

ALDOGAN, D.; YASLAN, Y. A comparison study on ensemble strategies and feature sets for sentiment analysis. *Lecture Notes in Electrical Engineering*, Elsevier Ltd, v. 363, p. 359–370, 2016. ISSN 18761119.

AMANCIO, D. R. et al. A Systematic Comparison of Supervised Classifiers. *PLoS ONE*, v. 9, n. 4, p. e94137, apr 2014. ISSN 1932-6203. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0094137>>.

ANTHONISSE, J. M. The rush in a directed graph. *Stichting Mathematisch Centrum. Mathematische Besliskunde*, BN 9/71, p. 10, 1971.

AZIZI, N.; FARAH, N. From static to dynamic ensemble of classifiers selection: Application to Arabic handwritten recognition. *International Journal of Knowledge-based and Intelligent Engineering Systems*, v. 16, n. 4, p. 279–288, dec 2012. ISSN 18758827. Disponível em: <<http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/KES-2012-00249>>.

BANFIELD, R. E. et al. Ensemble diversity measures and their application to thinning. *Information Fusion*, v. 6, n. 1, p. 49–62, 2005. ISSN 15662535.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 11, 1999. ISSN 00368075. Disponível em: <<http://arxiv.org/abs/cond-mat/9910332http://www.sciencemag.org/cgi/doi/10.1126/science.286.5439.509>>.

BARDDAL, J. P.; GOMES, H. M.; ENEMBRECK, F. SFNClassifier. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC '14*. New York, New York, USA: ACM Press, 2014. p. 786–791. ISBN 9781450324694. Disponível em: <<http://dl.acm.org/citation.cfm?id=2554855http://dl.acm.org/citation.cfm?doid=2554850.2554855>>.

BOLDI, P.; VIGNA, S. Axioms for centrality. *Internet Mathematics*, v. 10, n. 3-4, p. 222–262, 2014. ISSN 15427951.

BONACICH, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, v. 2, p. 113–120, 1972. Disponível em: <<http://ci.nii.ac.jp/naid/10016154506/en/>>.

BONACICH, P. Some unique properties of eigenvector centrality. *Social Networks*, v. 29, n. 4, p. 555–564, 2007. ISSN 03788733.

BORGATTI, S. P. Centrality and network flow. *Social Networks*, v. 27, n. 1, p. 55–71, 2005. ISSN 03788733.

BRANDES, U. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, v. 30, n. 2, p. 136–145, 2008. ISSN 03788733.

BREIMAN, L. Bagging Predictors. n. 421, 1994.

Britto Jr., A. d. S.; SABOURIN, R.; OLIVEIRA, L. E. S. Dynamic selection of classifiers - A comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665–3680, 2014. ISSN 00313203.

BROWN, G.; KUNCHEVA, L. I. "Good" and "bad" diversity in majority vote ensembles. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 5997 LNCS, p. 124–133, 2010. ISSN 03029743.

BUNTINE, W. L. *A Theory Of Learning Classification Rules*. Tese (Doutorado) — University of Technology, Sydney, 1992. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.5614>>.

BURDUK, R. Integration Base Classifiers Based on Their Decision Boundary. In: Rutkowski L., Korytkowski M., Scherer R., Tadeusiewicz R., Zadeh L., Z. J. (Ed.). *Artificial Intelligence and Soft Computing: 16th International Conference, ICAISC 2017, Zakopane, Poland, June 11-15, 2017, Proceedings, Part II*. Springer International Publishing, 2017. v. 10246, p. 13–20. ISBN 978-3-319-59062-2. Disponível em: <http://link.springer.com/10.1007/978-3-319-59063-9http://link.springer.com/10.1007/978-3-319-59060-8_2>.

CAVALCANTI, G. D. C. et al. Combining diversity measures for ensemble pruning. *Pattern Recognition Letters*, Elsevier B.V., v. 74, p. 38–45, 2016. ISSN 01678655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2016.01.029>>.

CHEN, D. et al. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 391, n. 4, p. 1777–1787, feb 2012. ISSN 03784371. Disponível em: <<http://dx.doi.org/10.1016/j.physa.2011.09.017http://linkinghub.elsevier.com/retrieve/pii/S0378437111007333>>.

CONTRACTOR, N.; FORBUSH, E. Networks. *The International Encyclopedia of Organizational Communication*, p. 1–25, 2017. Disponível em: <<http://doi.wiley.com/10.1002/9781118955567.wbieoc148>>.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. META-DES.Oracle: Meta-learning and feature selection for dynamic ensemble selection. *Information Fusion*, Elsevier B.V., v. 38, p. 84–103, 2017. ISSN 15662535.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, Elsevier B.V., v. 41, p. 195–216, may 2018. ISSN 15662535. Disponível em: <<http://dx.doi.org/10.1016/j.inffus.2017.09.010http://linkinghub.elsevier.com/retrieve/pii/S1566253517304074>>.

CRUZ, R. M. O.; SABOURIN, R.; CAVALCANTI, G. D. C. Analyzing different prototype selection techniques for dynamic classifier and ensemble selection. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017. p. 3959–3966. ISBN 978-1-5090-6182-2. Disponível em: <<http://ieeexplore.ieee.org/document/7966355/>>.

CUNNINGHAM, P.; CARNEY, J. Diversity versus quality in classification ensembles based on feature selection. *Proceedings of the 11th European Conference on Machine Learning*, p. 109–116, 2000. ISSN 16113349. Disponível em: <http://books.google.com/books?hl=en&lr=&id=M9WuI6tiqRcC&oi=fnd&pg=PA231&dq=Neural+network+ensembles,+cross+validation,+and+active+learning&ots=Gu4ueT3yQD&sig=58HkTCGwPyvEZsxJpv51LQGhsOs%5Cnhttp://link.springer.com/chapter/10.1007/3-540-45164-1_12>.

DAI, Q.; YE, R.; LIU, Z. Considering diversity and accuracy simultaneously for ensemble pruning. *Applied Soft Computing Journal*, Elsevier B.V., v. 58, p. 75–91, 2017. ISSN 15684946. Disponível em: <<http://dx.doi.org/10.1016/j.asoc.2017.04.058>>.

DEMŠAR, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *The Journal of Machine Learning Research*, v. 7, p. 1–30, 2006. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=1248547.1248548>>.

DIANATI, N. Unwinding the hairball graph: pruning algorithms for weighted complex networks. *Phys. Rev. E*, v. 93, n. 1, p. 8, 2016. ISSN 15502376. Disponível em: <<http://arxiv.org/abs/1503.04085>>.

DIETTERICH, T. G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, v. 40, p. 139–157, 2000. Disponível em: <<http://en.scientificcommons.org/42637098%5Cnuuid/7906280C-AEF8-405A-9A94-6BAA1DDAED1E>>.

DUIN, R. et al. *Prtools: a Matlab toolbox for pattern recognition*. 2017. Disponível em: <<http://www.prtools.org>>.

DUTOT, A. et al. GraphStream: A Tool for bridging the gap between Complex Systems and Dynamic Graphs. *Emergent Properties in Natural and Artificial Complex Systems. Satellite Conference within the 4th European Conference on Complex Systems (ECCS'2007)*, 2007. Disponível em: <<http://arxiv.org/abs/0803.2093http://hal.archives-ouvertes.fr/hal-00264043>>.

EK, B. et al. A comprehensive comparison of graph theory metrics for social networks. *Social Network Analysis and Mining*, Springer Vienna, v. 5, n. 1, p. 37, 2015. ISSN 1869-5450. Disponível em: <<http://link.springer.com/10.1007/s13278-015-0272-2>>.

ENEMBRECK, F.; BARTHÈS, J.-P. A. A social approach for learning agents. *Expert Systems with Applications*, v. 40, n. 5, p. 1902–1916, 2013. ISSN 09574174. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0957417412011220>>.

ERDÖS, P.; RÉNYI, a. On random graphs. *Publicationes Mathematicae*, v. 6, p. 290–297, 1959. ISSN 00029947.

FLEISS, J. L.; LEVIN, B.; PAIK, M. C. *Statistical Methods for Rates and Proportions*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003. 800 p. (Wiley

Series in Probability and Statistics). ISBN 0471526290. Disponível em: <<http://doi.wiley.com/10.1002/0471445428>>.

FRANK, E.; HALL, M. A.; WITTEN, I. H. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition*. 2016. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf>.

FREEMAN, L. C. Centrality in social networks conceptual clarification. *Social Networks*, v. 1, n. 3, p. 215–239, 1978. ISSN 03788733.

FREEMAN, L. C. The development of social network analysis. *A Study in the Sociology of Science*, Empirical Press, North Charleston, South Carolina, p. 218, 2004. ISSN 03788733. Disponível em: <http://www.researchgate.net/profile/Linton_Freeman/publication/239228599_The_Development_of_Social_Network_Analysis/links/54415c650cf2e6f0c0f616a8.pdf>.

GARGIULO, F.; MAZZARIELLO, C.; SANSONE, C. Multiple Classifier Systems: Theory, Applications and Tools. In: BIANCHINI, M.; MAGGINI, M.; JAIN, L. C. (Ed.). *Handbook on Neural Information Processing*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2013, (Intelligent Systems Reference Library, v. 49). cap. Chapter 10, p. 335–378. ISBN 978-3-642-36656-7. Disponível em: <http://link.springer.com/10.1007/978-3-642-36657-4http://link.springer.com/10.1007/978-3-642-36657-4_10>.

GIACINTO, G. Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing*, v. 19, p. 699–707, 2001. ISSN 02628856. Disponível em: <[http://dx.doi.org/10.1016/S0262-8856\(01\)00045-2](http://dx.doi.org/10.1016/S0262-8856(01)00045-2)>.

GIACINTO, G.; ROLI, F. An approach to the automatic design of multiple classifier systems. *Pattern Recogn. Lett.*, v. 22, n. 1, p. 25–33, 2001.

GOMES, H. M. et al. A Survey on Ensemble Learning for Data Stream Classification. *ACM Computing Surveys*, v. 50, n. 2, p. 1–36, 2017. ISSN 03600300. Disponível em: <<http://dl.acm.org/citation.cfm?doid=3071073.3054925>>.

GOMES, H. M.; ENEMBRECK, F. SAE2: Advances On The Social Adaptive Ensemble Classifier for Data Streams. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC '14*. New York, New York, USA: ACM Press, 2014. p. 798–804. ISBN 9781450324694. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2554850.2554905>>.

HAI XIANG, G. et al. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, Elsevier Ltd, v. 73, p. 220–239, 2017. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2016.12.035>>.

HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine*, v. 12, n. 10, p. 993–1001, 1990. ISSN 01628828. Disponível em: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Neural+Network+Ensembles#0%5Chttp://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=58871%5Cfile:///C:/Users/TUL/Documents/Litteratur/Hansen,Salamon-1990-Neuralnetworkensembles.pdf>.

- HO, T. K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 8, p. 832–844, 1998. ISSN 01628828.
- JACKSON, C. Using social network analysis to reveal unseen relationships in medieval Scotland. *Digital Scholarship in the Humanities*, v. 32, n. 2, p. 336–343, 2017. ISSN 2055768X.
- JAIN, A.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 4–37, 2000. ISSN 0162-8828. Disponível em: <http://ieeexplore.ieee.org/ielx5/34/17859/00824819.pdf?tp=&arnumber=824819&isnumber=17859%5Cnhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=824819>.
- JOHANSSON, U.; LÖFSTRÖM, T.; BOSTRÖM, H. Overproduce-and-Select : The Grim Reality. *Proceedings of the IEEE Symposium on Computational Intelligence and Ensemble Learning*, p. 52–59, 2013.
- JUANG, B. H.; KATAGIRI, S. Discriminative Learning for Minimum Error Classification. *IEEE Transactions on Signal Processing*, v. 40, n. 12, p. 3043–3054, 1992. ISSN 19410476.
- KAPP, M. N.; SABOURIN, R.; MAUPIN, P. An empirical study on diversity measures and margin theory for ensembles of classifiers. In: *2007 10th International Conference on Information Fusion*. IEEE, 2007. p. 1–8. ISBN 978-0-662-45804-3. Disponível em: <<http://ieeexplore.ieee.org/document/4408144/>>.
- KING, R. D.; FENG, C.; SUTHERLAND, A. Statlog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, v. 9, n. 3, p. 289–333, 1995. Disponível em: <<http://dx.doi.org/10.1080/08839519508945477>>.
- KITTLER, J. et al. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 3, p. 226–239, mar 1998. ISSN 01628828. Disponível em: <<http://ieeexplore.ieee.org/document/667881/>>.
- KO, A. H. R. et al. Pairwise fusion matrix for combining classifiers. *Pattern Recognition*, v. 40, n. 8, p. 2198–2210, 2007. ISSN 00313203.
- KOHAVI, R.; WOLPERT, D. H. Bias plus variance decomposition for zero-one loss functions. *Proceedings of the 13th International Conference on Machine Learning*, p. 275–283, 1996.
- KOTSIANTIS, S. B. Bagging and boosting variants for handling classifications problems: a survey. *The Knowledge Engineering Review*, v. 29, n. 01, p. 78–100, 2014. ISSN 0269-8889. Disponível em: <http://www.journals.cambridge.org/abstract_S0269888913000313>.
- KRAWCZYK, B. et al. Ensemble learning for data stream analysis: A survey. *Information Fusion*, v. 37, p. 132–156, 2017. ISSN 15662535.
- KRAWCZYK, B.; WOŹNIAK, M. Untrained weighted classifier combination with embedded ensemble pruning. *Neurocomputing*, v. 196, p. 14–22, 2016. ISSN 18728286.
- KUNCHEVA, L. *Ludmila Kuncheva Collection LKC*. 2004. Available: <http://pages.bangor.ac.uk/mas00a/activi> p. Disponível em: <http://pages.bangor.ac.uk/mas00a/activities/real_data.htm>.

- KUNCHEVA, L.; WHITAKER, C. Ten measures of diversity in classifier ensembles: Limits for two classifiers. *Proc. Inst. Elect. Eng. Workshop Intell. Sens. Process.*, p. 10–1–10–6, 2001. ISSN 09633308. Disponível em: <<http://hdl.handle.net/10242/41782>>.
- KUNCHEVA, L. I. *Combining Pattern Classifiers*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004. ISBN 0471210781. Disponível em: <<http://doi.wiley.com/10.1002/0471660264>>.
- KUNCHEVA, L. I. *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. [S.l.]: John Wiley & Sons, 2014, 2014. 1–357 p. ISSN 1045-9227. ISBN 9781118914564.
- 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. ISSN 08856125.
- KUNCHEVA, L. I. et al. Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis and Applications*, v. 6, n. 1, p. 22–31, 2003. ISSN 14337541.
- LATORA, V.; NICOSIA, V.; RUSSO, G. *Complex Networks*. Cambridge University Press, 2017. 594 p. ISBN 9781107103184. Disponível em: <<https://www.cambridge.org/core/product/identifier/9781316216002/type/book>>.
- LEWIS, T. G. *Network Science*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2009. ISBN 9780470400791. Disponível em: <<http://doi.wiley.com/10.1002/9780470400791>>.
- LI, N.; YU, Y.; ZHOU, Z. H. Diversity regularized ensemble pruning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 7523 LNAI, n. PART 1, p. 330–345, 2012. ISSN 03029743.
- LI, W.; HOU, J.; YIN, L. A classifier fusion method based on classifier accuracy. *2014 International Conference on Mechatronics and Control (ICMC)*, v. 32, n. Icmc, p. 2119–2122, 2014. ISSN 15662535. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1566253516300033><http://ieeexplore.ieee.org/document/7231940>/<http://ieeexplore.ieee.org/document/7231940>>.
- LI, Y. et al. A New Diversity Measure for Classifier Fusion. In: F.L., W. et al. (Ed.). *Multimedia and Signal Processing. Communications in Computer and Information Science*. Springer, Berlin, Heidelberg, 2012. v. 346, p. 396–403. ISBN 9783642352850. Disponível em: <http://link.springer.com/10.1007/978-3-642-35286-7_50>.
- LICHMAN, M. *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Sciences, 2013. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- LIU, Z. et al. Combination of classifiers with optimal weight based on evidential reasoning. *IEEE Transactions on Fuzzy Systems*, v. 6706, n. c, p. 1–1, 2017. ISSN 1063-6706. Disponível em: <<http://ieeexplore.ieee.org/document/7956193>>.
- LÜ, L. et al. Vital nodes identification in complex networks. *Physics Reports*, Elsevier B.V., v. 650, p. 1–63, 2016. ISSN 03701573. Disponível em: <<http://dx.doi.org/10.1016/j.physrep.2016.06.007>>.

- MA, H.; KING, I.; LYU, M. R. Learning to recommend with social trust ensemble. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '09*, p. 203, 2009. ISSN 2157-6904. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1571941.1571978>>.
- MAO, S. et al. Greedy optimization classifiers ensemble based on diversity. *Pattern Recognition*, Elsevier, v. 44, n. 6, p. 1245–1261, 2011. ISSN 00313203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2010.11.007>>.
- MARGINEANTU, D. D.; DIETTERICH, T. G. Pruning adaptive boosting. In: *ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997. p. 211–218.
- MARTÍNEZ-MUÑOZ, G.; SUÁREZ, A. Pruning in ordered bagging ensembles. *Proceedings of the 23rd international conference on Machine learning - ICML '06*, n. Icml, p. 609–616, 2006. ISSN 10987576. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1143844.1143921>>.
- MARTÍNEZ-MUÑOZ, G.; SUÁREZ, A. Out-of-bag estimation of the optimal sample size in bagging. *Pattern Recognition*, Elsevier, v. 43, n. 1, p. 143–152, jan 2010. ISSN 00313203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2009.05.010><http://linkinghub.elsevier.com/retrieve/pii/S003132030900212X>>.
- MARTÍNEZ-MUÑOZ, G. et al. Aggregation ordering in bagging. *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, p. 258–263, 2004.
- MORENO, J. L. *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Washington, D. C.: Beacon House, 1934. ISBN 9992695722.
- MORENO-SECO, F. et al. Comparison of classifier fusion methods for classification in pattern recognition tasks. In: YEUNG, D.-Y. et al. (Ed.). *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 705–713. ISBN 978-3-540-37241-7.
- NABIHA, A.; NADIR, F.; MOKHTAR, S. Progressive Algorithm for Classifier Ensemble Construction Based on Diversity : Application to the Arabic Handwritten Recognition. *The 2nd International Conference on Information and Communication Systems ICICS'11*, p. 27–33, 2011.
- NEWMAN, M. et al. *The Structure and dynamics of networks*. 2006. 419–421 p.
- NEWMAN, M. E. J. *Networks: An Introduction*. [S.l.]: OXFORD UNIVERSITY PRESS, 2010. 720 p. ISSN 18600832. ISBN 9783319035178.
- PARTALAS, I.; TSOUMAKAS, G.; VLAHAVAS, I. An ensemble uncertainty aware measure for directed hill climbing ensemble pruning. *Machine Learning*, v. 81, n. 3, p. 257–282, 2010. ISSN 08856125.
- PARTALAS, I.; TSOUMAKAS, G.; VLAHAVAS, I. A Study on Greedy Algorithms for Ensemble Pruning. 2012.

- PARTRIDGE, D.; KRZANOWSKI, W. Software diversity: practical statistics for its measurement and exploitation. *Information and Software Technology*, v. 39, n. 10, p. 707–717, 1997. ISSN 09505849.
- PARTRIDGE, D.; YATES, W. B. Engineering Multiversion Neural-Net Systems. *Neural Computation*, v. 8, p. 869–893, 1996.
- Ponti Jr., M. P. Combining Classifiers: From the Creation of Ensembles to the Decision Fusion. In: *2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*. IEEE, 2011. p. 1–10. ISBN 978-1-4577-1627-0. Disponível em: <<http://ieeexplore.ieee.org/document/6076744/>>.
- ROKACH, L. *Ensemble Methods for Classifiers*. New York: Springer-Verlag, 2010a. ISBN 9780387098227. Disponível em: <http://link.springer.com/10.1007/978-0-387-09823-4http://link.springer.com/10.1007/0-387-25465-X_45>.
- ROKACH, L. Ensemble-based classifiers. *Artificial Intelligence Review*, v. 33, n. 1, p. 1–39, Feb 2010b. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-009-9124-7>>.
- ROY, A. et al. A study on combining dynamic selection and data preprocessing for imbalance learning. *Neurocomputing*, Elsevier B.V., v. 286, n. February, p. 179–192, apr 2018. ISSN 09252312. Disponível em: <<https://doi.org/10.1016/j.neucom.2018.01.060http://linkinghub.elsevier.com/retrieve/pii/S0925231218300936>>.
- RUHNAU, B. Eigenvector-centrality — a node-centrality? *Social Networks*, v. 22, p. 357–365, 2000. ISSN 03788733.
- RUTA, D.; GABRYS, B. Classifier selection for majority voting. *Information Fusion*, v. 6, n. 1, p. 63–81, mar 2005. ISSN 15662535. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1566253504000417>>.
- SCHAPIRE, R. E. et al. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, v. 26, n. 5, p. 1651–1686, 1998. ISSN 00905364.
- SEATH, A.; SOKAL, R. Numerical Taxonomy. The Principles and Practice of Numerical Classification. *Systematic Zoology*, W H Freeman & Co (Sd) (June 1973), v. 24, n. 2, p. 263–268, 1973. ISSN 00397989.
- SILVA, R. A. et al. Fusion of Classifiers based on Centrality Measures. In: *30th International Conference on Tools with Artificial Intelligence*. [S.l.: s.n.], 2018. p. 1–8. No prelo.
- SINGH, P. K.; SARKAR, R.; NASIPURI, M. Correlation-based classifier combination in the field of pattern recognition. *Computational Intelligence*, n. September 2016, p. 1–36, 2017. ISSN 14678640.
- SKALAK, D. The sources of increased accuracy for two proposed boosting algorithms. *Proc. American Association for Artificial Intelligence, . . .*, p. 120—125, 1996. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.2269&rep=rep1&type=pdf%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+sources+of+increased+accuracy+for+two+proposed+boosting+algorithms#0>>.

- SOSNOWSKA, J.; SKIBSKI, O. Attachment Centrality for Weighted Graphs. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, p. 416–422, 2017. Disponível em: <<https://www.ijcai.org/proceedings/2017/59>>.
- TRAWINSKI, K.; CORDON, O. A network-based approach for diversity visualization of fuzzy classifier ensembles. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2016. p. 923–930. ISBN 978-1-5090-0626-7. Disponível em: <<http://ieeexplore.ieee.org/document/7737787/>>.
- TSYMBAL, A.; PECHENIZKIY, M.; CUNNINGHAM, P. Diversity in Ensemble Feature Selection. *Department of Computer Science, Trinity College Dublin*, p. 1–38, 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.2105>>.
- TSYMBAL, A.; PUURONEN, S.; PATTERSON, D. W. Ensemble feature selection with the simple Bayesian classification. *Information Fusion*, v. 4, n. 2, p. 87–100, 2003. ISSN 15662535.
- ULAS, C. et al. A Supervised Learning Approach for the Fusion of Multiple Classifier Outputs. *International Journal of Signal Processing Systems*, v. 4, n. 3, p. 198–203, 2016.
- VALENTE, T. W. et al. How correlated are network centrality measures? *Connections*, v. 28, n. 1, p. 16–26, 2008. ISSN 0226-1766.
- VALENTINI, G. An experimental bias-variance analysis of SVM ensembles based on resampling techniques. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, v. 35, n. 6, p. 1252–1271, 2005. ISSN 10834419.
- VALENTINI, G.; MASULLI, F. Ensembles of Learning Machines. *Neural Nets: 13th Italian Workshop on Neural Nets, WIRN VIETRI 2002, Vietri sul Mare, Italy, May 30–June 1, 2002, Revised Papers*, v. 2486, p. 3–20, 2002. ISSN 03029743.
- WANG, S.; MINKU, L. L.; YAO, X. A systematic study of online class imbalance learning with concept drift. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2018.
- WASSERMAN, S.; FAUST, K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994. ISBN 0521387078. Disponível em: <[citeulike-article-id:165614%5Cnhttp://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0521387078](http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0521387078)>.
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. June, p. 440–442, 2008.
- WOZNIAK, M. *Hybrid Classifiers: Methods of Data, Knowledge, and Classifier Combination*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. v. 519. 217 p. (Studies in Computational Intelligence, v. 519). ISSN 1860949X. ISBN 978-3-642-40996-7. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-84885105209&partnerID=tZOtx3y1http://link.springer.com/10.1007/978-3-642-40997-4>>.
- WOŹNIAK, M.; GRAÑA, M.; CORCHADO, E. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, v. 16, n. 1, p. 3–17, 2014. ISSN 15662535.

- YIN, X. C. et al. A novel classifier ensemble method with sparsity and diversity. *Neurocomputing*, Elsevier, v. 134, p. 214–221, 2014. ISSN 18728286. Disponível em: <<http://dx.doi.org/10.1016/j.neucom.2013.07.054>>.
- YULE, G. U. On the Association of Attributes in Statistics: With Illustrations from the Material of the Childhood Society. *Philosophical Transactions of the Royal Society of London. 194(A)*; A, n. 194, p. 257–319, 1990.
- ZHOU, F. *Methods for network abstraction*. [s.n.], 2012. ISBN 9789521081576. Disponível em: <<https://helda.helsinki.fi/handle/10138/34840>>.
- ZHOU, F.; MAHLER, S.; TOIVONEN, H. Simplification of networks by edge pruning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 7250, n. Icdm, p. 179–198, 2012. ISSN 03029743.
- ZHOU, Z.-H. *Ensemble Methods: Foundations and Algorithms*. CRC Press, 2012. v. 1st. 236 p. ISSN 9781439830031. ISBN 1439830037. Disponível em: <<http://doi.wiley.com/10.1111/insr.12042><http://ieeexplore.ieee.org/document/6410720/>>.
- ZWEIG, K. A. *Network Analysis Literacy*. Vienna: Springer Vienna, 2016. 535 p. (Lecture Notes in Social Networks). ISBN 978-3-7091-0740-9. Disponível em: <<http://www.springer.com/series/8768>http://dx.doi.org/10.1007/978-3-319-04147-6_11<http://link.springer.com/10.1007/978-3-7091-0741-6>>.

Apêndices

APÊNDICE A – CBS com Bagging

Neste apêndice, são descritas as análises realizadas para o método proposto de seleção estática, CBS, com a literatura. O método de geração de *pool* adotado foi o Bagging, enquanto que na Seção 5.2 o método foi o Boosting. Os demais parâmetros usados como o tamanho do *ensemble*, o classificador usado e demais ferramentas foram as mesmas que os apresentados para o CBF. Portanto, o método de geração foi o Bagging, com tamanho de bag reduzido a 66% do tamanho da base de treino S_{train} , isto é, cada base de dados obtida por reamostragem com reposição corresponde a 2/3 do tamanho de S_{train} . Foram gerados 100 diferentes classificadores a partir do treinamento do algoritmo Perceptron de Erro Quadrado Mínimo (*Perceptron with Minimum Square Error*) (JUANG; KATAGIRI, 1992) em cada uma das bases de treinamento geradas na reamostragem Bagging.

Na Tabela 12 é apresentada a comparação de CBS com a literatura, apresentando a média, o desvio-padrão, a quantidade de melhores resultados ($\#Best$) e o teste estatístico de Wilcoxon (WS). Nota-se que a abordagem proposta não apresenta a maior parte dos melhores resultados observados, apesar de ser uma abordagem competitiva em relação a DREP, que apresentou a melhor quantidade neste quesito. O teste estatístico de Wilcoxon foi avaliado e também mostrou que não há diferença significativa entre CBS e cada um dos métodos da literatura que foram comparados.

Tabela 12 – Comparação de CBS com métodos da literatura. O *pool* de classificadores foi gerado por Bagging. Os valores apresentados são a taxa de reconhecimento obtida por k-fold com $k = 6$. O valor ao lado ($|C'|$) refere-se ao tamanho médio do subconjunto selecionado. Em negrito está o melhor resultado daquele problema de classificação. # Best é quantidade de melhores resultados atingidas por um método. WS significa teste de sinais de Wilcoxon, sendo que o valor representado é referente ao valor de p , enquanto que (+) é para sinalizar resultados significativos com $\alpha = 0.05$.

	Kappa	$ C' $	DREP	$ C' $	POBE	$ C' $	AGOB	$ C' $	SB	$ C' $	MV	$ C' $	CBS	$ C' $
Australian	87.25 ± 2.87	50	87.83 ± 2.48	4	86.65 ± 1.96	20	86.65 ± 2.43	30	85.66 ± 3.05	1	84.21 ± 2.84	100	85.08 ± 2.16	10
Banana	84.35 ± 1.17	50	84.80 ± 1.04	18	84.35 ± 0.73	99	84.50 ± 0.98	29	84.50 ± 0.65	1	84.70 ± 1.12	100	83.35 ± 1.11	38
Blood	77.58 ± 2.45	50	77.96 ± 2.75	6	77.58 ± 1.35	34	76.38 ± 2.22	31	76.88 ± 3.47	1	77.55 ± 2.52	100	77.55 ± 3.77	11
CTG	88.52 ± 1.32	50	89.05 ± 0.92	3	88.52 ± 0.74	19	89.92 ± 0.96	28	88.48 ± 0.34	1	89.32 ± 1.16	100	89.32 ± 1.14	7
Diabetes	75.85 ± 3.31	50	77.16 ± 2.78	5	75.25 ± 2.25	28	76.65 ± 2.77	31	76.25 ± 4.67	1	76.12 ± 2.91	100	76.11 ± 2.38	8
Ecoli	85.42 ± 2.81	50	84.23 ± 3.49	5	85.42 ± 2.13	26	86.72 ± 2.92	30	83.33 ± 4.33	1	86.01 ± 2.99	100	84.82 ± 3.05	17
Faults	69.19 ± 2.78	50	68.99 ± 2.15	3	60.99 ± 1.61	15	68.99 ± 2.13	30	69.81 ± 1.91	1	66.79 ± 2.98	100	69.35 ± 2.55	5
German	76.00 ± 2.58	50	76.20 ± 2.34	4	74.20 ± 1.66	24	76.00 ± 2.19	31	72.50 ± 2.08	1	71.30 ± 3.57	100	75.40 ± 2.42	6
Glass	65.36 ± 11.42	50	67.22 ± 8.46	3	64.56 ± 7.05	18	67.22 ± 8.77	37	57.90 ± 6.35	1	59.78 ± 12.60	100	59.30 ± 8.68	5
Haberman	76.47 ± 4.18	50	77.78 ± 5.66	23	75.47 ± 3.68	65	73.97 ± 4.76	33	73.86 ± 7.82	1	76.47 ± 5.43	100	73.53 ± 3.71	32
Heart	81.85 ± 7.85	50	82.96 ± 8.51	5	82.05 ± 6.20	13	81.95 ± 7.74	34	80.37 ± 7.53	1	82.22 ± 6.24	100	81.85 ± 5.94	16
ILPD	71.01 ± 1.71	50	71.19 ± 2.51	4	70.11 ± 1.10	10	70.51 ± 1.87	29	71.19 ± 2.34	1	71.53 ± 2.50	100	72.56 ± 3.04	10
Ionosphere	80.91 ± 1.80	50	81.20 ± 2.20	2	83.64 ± 1.14	6	81.20 ± 1.77	30	84.33 ± 1.58	1	85.76 ± 2.88	100	81.20 ± 2.59	2
Laryngeal1	81.18 ± 6.88	50	82.58 ± 7.05	4	78.88 ± 5.02	12	81.18 ± 6.44	34	83.06 ± 5.71	1	78.58 ± 5.48	100	82.13 ± 5.84	8
Laryngeal3	70.85 ± 5.20	50	71.68 ± 2.08	6	69.65 ± 2.33	34	70.25 ± 2.87	32	70.27 ± 3.68	1	71.98 ± 4.12	100	72.25 ± 3.93	6
Lithuanian	84.80 ± 1.83	50	84.25 ± 1.61	4	85.40 ± 1.12	22	83.60 ± 1.51	32	82.20 ± 2.47	1	83.45 ± 1.80	100	84.30 ± 1.81	4
Liver	69.33 ± 5.83	50	70.19 ± 6.22	4	69.13 ± 4.81	21	68.63 ± 5.78	35	66.37 ± 2.26	1	68.15 ± 5.74	100	69.88 ± 3.65	5
Magic	79.88 ± 0.34	50	79.63 ± 0.32	6	79.78 ± 0.21	31	80.68 ± 0.29	30	79.43 ± 0.27	1	79.43 ± 0.29	100	79.68 ± 0.37	7
Mammo	82.76 ± 3.25	50	82.17 ± 1.87	3	82.36 ± 1.66	16	82.56 ± 2.12	28	83.73 ± 1.44	1	83.01 ± 2.39	100	83.98 ± 2.69	21
Monk	80.56 ± 3.49	50	86.11 ± 4.88	3	82.86 ± 3.04	14	77.76 ± 4.03	30	87.50 ± 3.31	1	82.87 ± 3.73	100	86.34 ± 3.44	38
Phoneme	77.31 ± 1.09	50	77.98 ± 1.11	6	69.71 ± 0.71	35	80.31 ± 0.98	31	77.74 ± 1.07	1	77.42 ± 1.03	100	77.91 ± 1.17	7
Segmentation	92.73 ± 0.67	50	92.77 ± 0.57	4	91.67 ± 0.54	23	94.13 ± 0.60	28	92.77 ± 0.66	1	92.51 ± 0.59	100	93.34 ± 0.25	11
Sonar	80.52 ± 3.50	50	73.53 ± 4.37	7	79.65 ± 2.52	36	79.72 ± 3.60	33	70.14 ± 7.41	1	75.00 ± 3.83	100	80.31 ± 4.25	10
Thyroid	95.66 ± 2.01	50	95.66 ± 1.59	5	93.36 ± 1.26	27	95.16 ± 1.60	28	96.68 ± 1.45	1	95.80 ± 1.38	100	96.38 ± 1.63	17
Vehicle	76.32 ± 4.14	50	77.90 ± 2.67	4	75.12 ± 2.26	23	75.62 ± 2.89	33	77.42 ± 4.84	1	76.60 ± 3.46	100	76.83 ± 3.45	4
Vertebral	86.33 ± 4.82	50	86.33 ± 5.22	3	86.23 ± 3.60	17	85.13 ± 4.66	32	87.00 ± 4.86	1	86.33 ± 4.38	100	85.33 ± 4.27	13
WBC	96.84 ± 1.61	50	95.60 ± 1.89	6	97.04 ± 1.27	35	96.24 ± 1.65	28	96.48 ± 1.59	1	96.66 ± 1.54	100	97.19 ± 1.45	49
WDVG	86.56 ± 0.89	50	85.98 ± 1.06	3	87.36 ± 0.65	16	85.01 ± 0.90	28	86.26 ± 1.11	1	86.70 ± 1.11	100	86.36 ± 0.97	7
Weaning	80.80 ± 6.92	50	81.14 ± 3.28	4	81.00 ± 2.59	24	79.90 ± 3.84	33	80.14 ± 7.49	1	79.48 ± 7.37	100	80.45 ± 7.53	5
Wine	97.76 ± 2.49	50	97.78 ± 1.58	75	97.76 ± 1.27	46	97.76 ± 1.69	27	96.63 ± 2.77	1	97.20 ± 2.30	100	97.20 ± 2.30	97
AVRG	81.33 ± 3.37	50	81.59 ± 3.09	8	80.52 ± 2.21	27	81.14 ± 2.90	30	80.63 ± 3.28	1	80.76 ± 3.34	100	81.31 ± 3.05	16
# Best	1		13		2		6		5		1		4	
WS	0.71138		0.41222		0.17702		0.42952		0.11642		0.1031		N/A	

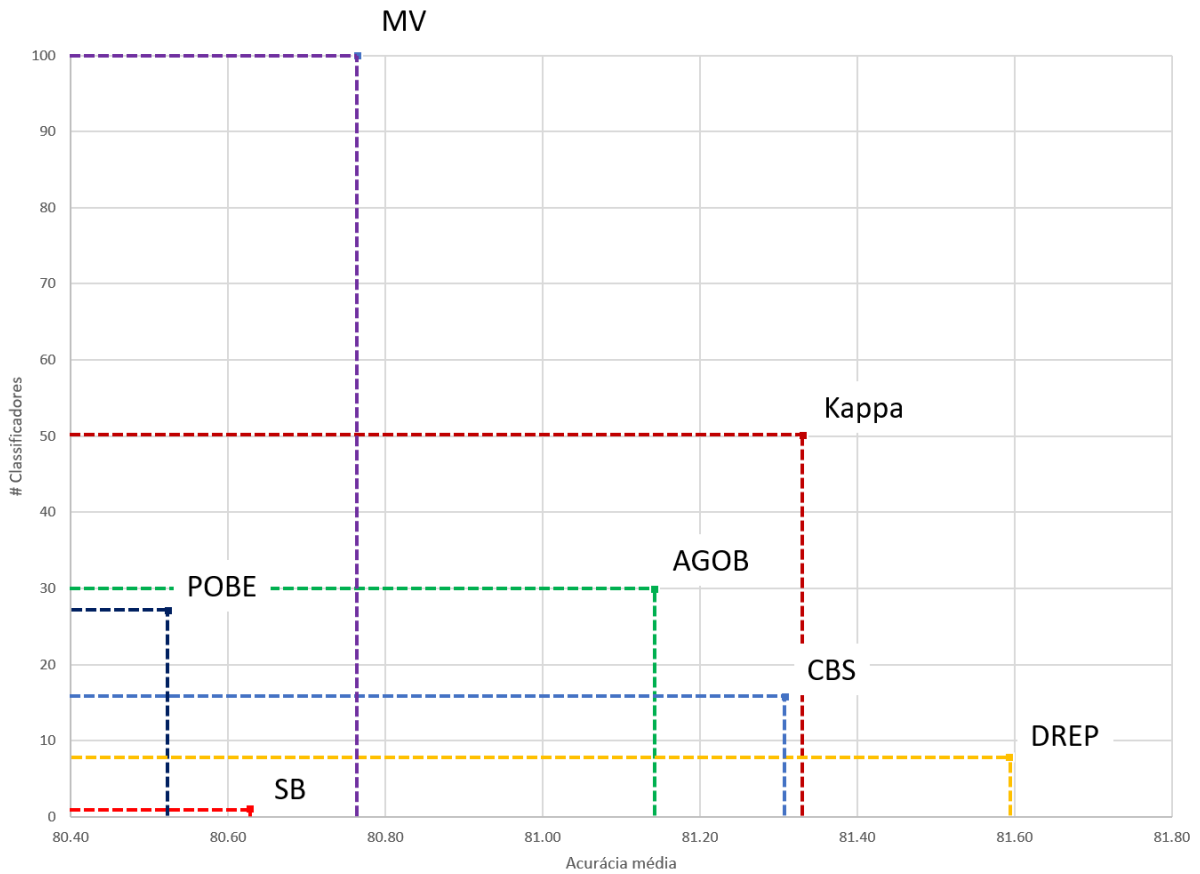


Figura 21 – Comparação entre abordagens de seleção quanto a acurácia média e o tamanho médio do conjunto de classificadores gerados por Bagging.

A Figura 21 apresenta a comparação entre os métodos quanto a quantidade de classificadores e a acurácia média. Neste caso, a abordagem CBS é inferior a DREP tanto na quantidade de classificadores (usa mais classificadores) quanto na acurácia média (menos acurado). Embora competitivo, o método CBS não apresentou melhor acurácia média que Kappa, entretanto, o tamanho de pool apresentado é bastante reduzido.

O teste post hoc de Nemenyi é apresentado na Figura 22. Percebe-se que os métodos são considerados semelhantes, uma vez que a distância crítica (CD) abrange a posição média de todas as abordagens. Outra observação é a de que a abordagem DREP possui uma posição média ligeiramente melhor que CBS. A posição é definida para cada problema de classificação, atribuindo 1 para a melhor solução (em média), 2 para a segunda melhor e assim por diante. A posição média é obtida pela média da posição de cada solução considerando os 30 problemas de classificação definidos no protocolo.

A Figura 23 apresenta a comparação pareada entre CBS e os métodos de seleção da literatura de maneira pareada. Percebe-se que quando o subconjunto de classificadores C' é selecionado pela abordagem CBS, apresenta acurácia interessante e competitiva

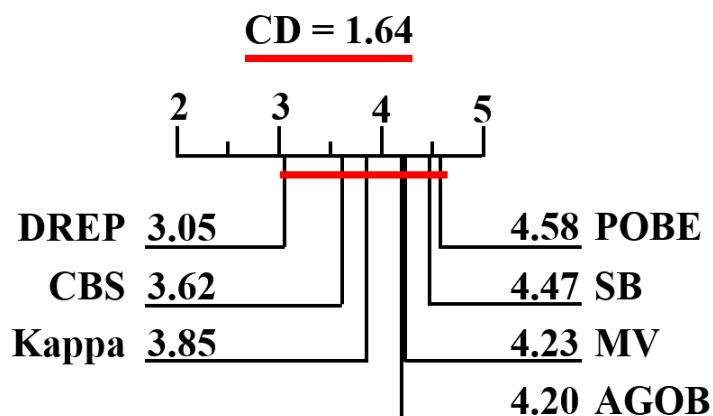


Figura 22 – Teste de Friedman e post hoc Nemenyi comparando CBS com a Literatura. O *pool* foi gerado por Bagging.

quando comparado a literatura, entretanto o resultado não é considerado significativo, pois a quantidade de vitórias para os 30 problemas de classificação não atinge o mínimo estabelecido pelo valor crítico ($cv = 19.5$), o qual é representado na figura pela linha horizontal.

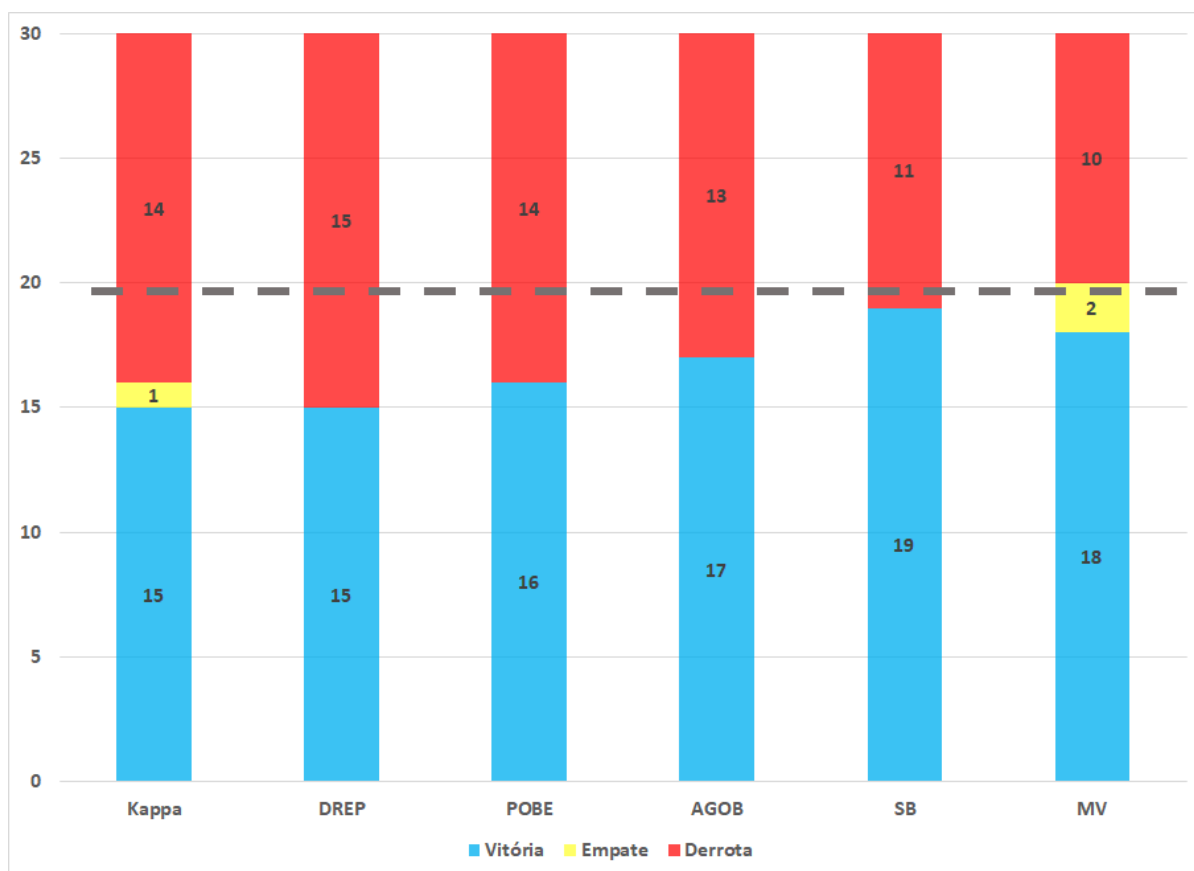


Figura 23 – Comparação de CBS com a literatura quando o pool é gerado por Bagging.

A análise dos experimentos sugere que o método CBS, apesar de competitivo,

demonstra resultado inferior a abordagem DREP quando o método de geração de classificadores é o Bagging. O método DREP apresenta algumas vantagens em relação a CBS, tais como uma posição média menor (Figura 22), usa menos classificadores em média e apresenta acurácia superior (Figura 21), além disso, apresenta a maior quantidade de melhores resultados (Tabela 12). Entretanto, DREP é semelhante a CBS segundo o critério da comparação pareada, assim como o teste de sinais de Wilcoxon aponta que os métodos são semelhantes estatisticamente. Além disso, o teste post hoc de Nemenyi aponta que esses métodos são semelhantes devido a distância crítica não ser suficiente alta.