

6

Experimentos com Tarefas de Classificação de Dados

De forma a validar o desempenho das abordagens BAS mostradas, foram conduzidos diversos experimentos com tarefas de classificação de dados. A maior preocupação desses experimentos é a de comparar tais abordagens com o algoritmo original AdaBoost. Em tais experimentos, não são conduzidas comparações com outros algoritmos de aprendizado de máquina, pois para tais tarefas existem diversos trabalhos (Sch00) comparando a eficiência das abordagens Boosting com outros métodos.

Nesse capítulo, são apresentadas as configurações BAS genéricas utilizadas nas tarefas de classificação de dados, bem como os resultados dos experimentos com as tarefas escolhidas.

6.1

Medidas de Qualidade

A medida principal de avaliação do desempenho de classificadores de tal natureza é a acurácia. A acurácia é definida como o número total de exemplos corretamente classificados pelo classificador dividido pelo número total de exemplos classificados. Portanto, temos

$$\text{acurácia} = \frac{|\{\text{exemplos corretamente classificados}\}|}{|\{\text{exemplos}\}|}$$

Nos experimentos com abordagens supervisionadas, como não foram fornecidas versões de treinamento e teste para cada conjunto de dados, aplicamos um esquema de validação cruzada chamado K-dobras (K-fold).

No K-dobras, o conjunto original é dividido em K subconjuntos. Em cada iteração k , um dos subconjuntos é retirado do conjunto de treinamento para testar o classificador gerado utilizando os outros $K - 1$ subconjuntos. Esse processo é repetido K vezes. Comumente, o valor de K utilizado é igual a dez.

Por outro lado, nos experimentos da abordagem semi-supervisionada cada conjunto é previamente dividido em K subconjuntos de treinamento e teste, permitindo que a validação cruzada seja realizada sobre tais subconjuntos.

Para cada experimento, são reportados a média e o desvio-padrão da acurácia nas K iterações de validação cruzada realizadas.

Tabela 6.1: Exemplos das instâncias do conjunto de dados *Weather*.

Atributos				Classe Alvo
perspectiva	temperatura	umidade	vento	jogar
ensolarado	quente	alta	não	não
ensolarado	quente	alta	sim	não
nublado	quente	alta	não	sim
chuvoso	ameno	alta	não	sim

6.2

Modelagem BAS para Classificação de Dados

Boosting normalmente é aplicado em tarefas de classificação. Em tais tarefas, o objetivo é prever o valor da classe alvo de um exemplo utilizando para isso os valores de seus atributos conhecidos.

Por exemplo, na tabela 6.1 é apresentado o conjunto de dados *Weather* descrito neste capítulo.

Em todos os experimentos deste capítulo, foram utilizados os mesmos parâmetros de Boosting. Tais parâmetros foram obtidos empiricamente para alguns conjuntos específicos e então aplicados para todos os conjuntos. O objetivo de utilizar os mesmos parâmetros para todas as tarefas é o de verificar a robustez das abordagens utilizadas aplicadas em problemas sobre os que não são se possui nenhum conhecimento.

A seguir os diversos parâmetros dos algoritmos derivados do BAS são listados.

Classificador Base: utilizamos o algoritmo Toco de Decisão, próprio para utilização em estratégias de Boosting, descrito na seção 5.1.

Número de Classificadores BAS: treinamos nove classificadores para formar o comitê BAS. Para alguns conjuntos foram treinados mais classificadores, inclusive, todos os possíveis, dado uma distribuição de pesos fixa. Entretanto, como não houve ganho significativo de desempenho, optamos por tal quantidade com o objetivo de minimizar o tempo de treinamento.

Tamanho do Conjunto de Validação: selecionamos vinte por cento do conjunto de dados para ser utilizado como conjunto de validação.

Algoritmo de agrupamento: Devido a sua simplicidade e bons resultados, aplicamos o algoritmo K-Médias.

Quantidade de grupos de exemplos: é permitido ao algoritmo de agrupamento determinar cinco grupos para os exemplos. Realizamos tes-

tes também com dois grupos, visto que os problemas são de classificação binária e com mais grupos. Entretanto os seus desempenhos não foram satisfatórios.

Distribuição para os pesos: utilizamos distribuições que não geram valores muito distantes para os pesos dos exemplos. Podemos citar, com esta característica, as distribuições obtidas por meios das progressões Aritmética e Exponencial e da lei de Zipf (Man99). Também utilizamos um método de ponderação híbrido, onde a cada iteração, utilizamos uma das distribuições citadas. Com essa configuração, os pesos podem ser iniciados para quaisquer um dos valores listados na Tabela 6.2, onde cada linha representa um possível Método de Ponderação (MP) para o Comitê BAS.

Tabela 6.2: Possíveis pesos iniciais para o algoritmo Comitê BAS.

MP	Nome	Distribuição de Pesos				
1	Aritmética	1	2	3	4	5
2	Geométrica	1	2	4	8	16
3	Zipf	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$
4	Híbrido	Uma das distribuições acima				

Esquema de Votação: foram testados os dois esquemas de votação propostos

1. Votação pela Melhor Maioria: nesse esquema de votação são escolhidos os três melhores classificadores BAS treinados. Constatamos que a utilização de um comitê com três membros possui melhor performance que qualquer um dos classificadores tomados isoladamente. Além disso, a escolha de cinco ou mais classificadores, na maioria dos casos, degradou a performance do comitê final, devido ao baixo desempenho individual de alguns dos classificadores gerados.
2. Votação ETL: aqui o algoritmo ETL é utilizado para determinar a melhor combinação entre os nove classificadores treinados.

Parâmetros Genéticos: No caso especial do algoritmo BAS Genético, utilizamos os seguintes parâmetros, comumente empregados em abordagens evolutivas e que possuem boa relação desempenho x tempo de treinamento.

Tamanho da população: 100.

Número de Gerações: 10.

Percentual de mutação: 2%.

Percentual de cruzamento: 80%.

6.3

Descrição dos Conjuntos de Dados

Para a realização dos experimentos com tarefas de classificação de dados, selecionamos alguns conjuntos de dados referenciados em outros artigos e disponíveis livremente na internet. Nesta seção, apresentamos uma descrição sucinta dos conjuntos utilizados.

6.3.1

Conjunto do Repositório UCI

De forma a validar o desempenho das estratégias BAS supervisionadas, escolhemos 19 conjuntos de dados extraídos do Repositório de aprendizado de máquina UCI (Uci98). Selecionamos tais conjuntos segundo alguns critérios. Primeiramente, a classe-alvo deveria ser do tipo categórica. Em segundo lugar, a classe-alvo deveria ser binária, ou pelo menos facilmente tornada binária, segundo algum critério de agrupamento de classes. Não restringimos os conjuntos de dados quanto a sua origem, nem ao tipo dos atributos. A Tabela 6.3 mostra a descrição de tais conjuntos.

Tabela 6.3: Conjuntos de dados utilizados nos experimentos supervisionados.

Nome do Conjunto	Abr.	Número de Exemplos	Número de Atributos	Distribuição das Classes
Audiology Standardized	aud	226	69	25 / 201
Breast Cancer	bca	286	9	201 / 85
Car Evaluation	car	1728	6	1594 / 134
Congressional Voting Records	vot	435	16	168 / 267
Contraceptive Method Choice	cmc	1473	9	629 / 844
Credit Screening	crx	690	15	307 / 383
Echocardiogram	ech	131	9	88 / 43
Glass Identification	gla	214	10	163 / 51
Haberman's Survival	hab	306	3	225 / 81
Hepatitis	hep	155	19	32 / 123
Horse Colic	hco	366	26	214 / 152
Lenses	len	24	4	15 / 9
Monks-2	mk2	601	6	395 / 206
Monks-3	mk3	554	6	266 / 288
Promoters	pro	106	60	53 / 53
Spect-Heart	spe	267	22	55 / 212
Tic-Tac-Toe	ttt	958	9	626 / 332
Titanic	tit	2201	3	1490 / 711
Weather	wea	14	4	5 / 9

Nos experimentos, utilizamos apenas tarefas envolvendo problemas de classificação binária. Em consequência disso, alguns conjuntos de dados são

adaptados de forma a concatenar algumas classes segundo um critério de refinamento. Tais concatenações estão indicadas na Tabela 6.4.

Tabela 6.4: Concatenação das classes para tarefas de classificação não-binária.

Abr.	Classes	
aud	mixed-cochlear-age-fixation	cochlear-unknown poss-central
car	unacc acc	vgood good
cmc	No-use	Short-term Long-term
gla	building-windows-float-processed building-windows-non-float processed vehicle-windows-float-processed	containers tableware headlamps
len	hard contact lenses soft contact lenses	no-contact-lenses

6.3.2

Conjunto do Repositório IDA

Para validar a estratégia semi-supervisionada do Comitê BAS, conduzimos experimentos com o esquema proposto utilizando 13 conjuntos de dados reais e artificiais (Ida98) comumente utilizados em diversos trabalhos relacionados a aprendizado semi-supervisionado e Boosting. A Tabela 6.5 mostra a descrição de tais conjuntos.

Tabela 6.5: Conjuntos de dados utilizados nos experimentos semi-supervisionados.

Nome do Conjunto	Abr.	Quantidade		Tamanho dos Conjuntos		
		Subconjuntos	Atributos	Treino	Teste	Amostras
Banana	ban	100	2	400	4900	39600
Breast-Cancer	bca	100	9	200	77	19800
Diabetes	dia	100	8	468	300	46332
Flare-Solar	fls	100	9	666	400	65934
German	ger	100	20	700	300	69300
Heart	hea	100	13	170	100	16830
Image	ima	20	18	1300	1010	24700
Ringnorm	rin	100	20	400	7000	39600
Splice	spl	20	60	1000	2175	19000
Titanic	tit	100	3	150	2051	14850
Thyroid	thy	100	5	140	75	13860
Twonorm	two	100	20	400	7000	39600
Waveform	wav	100	21	400	4600	39600

Cada um destes conjuntos é previamente dividido em diversos subconjuntos de treinamento e teste. Assim, o aprendizado semi-supervisionado é re-

alizado considerando os demais subconjuntos de treinamento, não envolvidos na iteração corrente, como amostras.

6.4

Software e Hardware

Todo o código-fonte envolvido na experimentação foi desenvolvido no Laboratório de Engenharia de Algoritmos e Redes Neurais (LEARN) da PUC-Rio. A linguagem utilizada para codificação dos algoritmos foi o Python (Lut96) versão 2.4. Os experimentos com tarefas de classificação de dados foram executados em um computador portátil com processador Intel® Core2Duo de 2.00 GHz e 3 GB de memória RAM.

6.5

Resultados

Nesta seção, apresentamos os resultados dos experimentos conduzidos utilizando tarefas de classificação de dados.

6.5.1

Experimentos com Abordagens Supervisionadas

Primeiramente, são apresentados os resultados com a abordagem genética de determinação da melhor distribuição inicial de pesos aplicada a apenas um classificador BAS. Na Tabela 6.6 apresentamos a acurácia média e o seu desvio-padrão sobre as dez iterações de validação cruzada executadas para cada instância, bem como o número total de classificadores-base empregados.

A seguir, apresentamos os resultados dos experimentos conduzidos nas mesmas tarefas, utilizando a estratégia Comitê BAS com um esquema final de votação pela melhor maioria. Na Tabela 6.7, é apresentada uma comparação entre as distribuições iniciais de pesos utilizadas, cada qual representando um método de ponderação (MP) diferente. Os possíveis métodos de ponderação são: Aritmética, Geométrica, Zipf e Híbrido.

Podemos perceber através da Tabela 6.7, onde os melhores resultados, em termos de comparação das acurácias médias, para cada instância estão em negrito, que a ponderação híbrida apresenta os melhores resultados na grande maioria dos casos. Por esse motivo e por permitir uma maior variação nos classificadores gerados, tal modo de ponderação foi utilizado nos experimentos com votação pela melhor maioria que se seguem.

Na Tabela 6.8, apresentamos os resultados do algoritmo Comitê BAS, aplicando o esquema de combinação final por Votação ETL. Nesse esquema, trocamos a votação pela melhor maioria no conjunto de validação por um

Tabela 6.6: Resultados para a abordagem BAS genético.

Instância	#Cl.	Acurácia	
aud	19	97.27	± 0.50
bca	16	74.29	± 0.20
car	80	97.85	± 0.05
cmc	70	68.16	± 0.15
crx	64	86.23	± 1.81
ech	24	92.31	± 2.13
gla	1	99.52	± 1.26
hab	10	75.67	± 0.65
hco	25	83.06	± 0.69
hep	33	84.67	± 1.30
len	11	91.67	± 2.12
mk2	9	65.83	± 2.55
mk3	18	97.09	± 0.70
pro	24	94.00	± 1.57
spe	26	83.08	± 1.49
tit	10	77.82	± 0.35
ttt	171	95.79	± 0.59
vot	24	97.21	± 0.52
wea	10	78.57	± 5.03

Tabela 6.7: Resultados para a abordagem Comitê BAS com votação pela melhor maioria.

Instância	Aritmética		Geométrica		Zipf		Híbrida	
	#Cl.	Acurácia	#Cl.	Acurácia	#Cl.	Acurácia	#Cl.	Acurácia
aud	61	97.27 ± 0.75	79	97.73 ± 0.63	65	95.91 ± 0.12	60	97.73 ± 0.63
bca	40	72.86 ± 0.60	68	75.36 ± 1.09	29	73.21 ± 3.47	79	73.57 ± 2.58
car	199	97.33 ± 0.42	220	96.16 ± 0.94	108	96.80 ± 0.08	144	97.21 ± 0.13
cmc	152	68.44 ± 0.11	185	67.69 ± 0.47	61	67.35 ± 2.46	157	68.71 ± 0.00
crx	139	86.09 ± 0.24	93	85.80 ± 1.53	30	85.80 ± 2.09	62	86.52 ± 0.12
ech	56	92.31 ± 2.13	64	93.08 ± 1.92	38	92.31 ± 2.13	56	93.08 ± 2.35
gla	3	99.52 ± 1.26	3	99.52 ± 1.26	3	99.05 ± 1.67	3	99.52 ± 1.26
hab	29	75.00 ± 2.32	55	75.00 ± 2.32	15	75.33 ± 0.55	39	77.00 ± 1.02
hco	50	83.33 ± 0.00	75	83.06 ± 1.62	21	82.50 ± 3.63	48	84.17 ± 1.00
hep	57	84.67 ± 0.55	70	86.00 ± 0.18	42	82.67 ± 1.11	63	84.00 ± 2.59
len	18	87.50 ± 3.17	24	87.50 ± 22.17	9	91.67 ± 10.56	17	91.67 ± 10.56
mk2	20	65.67 ± 3.98	23	65.83 ± 2.20	14	65.67 ± 4.35	24	65.83 ± 0.98
mk3	34	97.27 ± 0.75	36	97.09 ± 0.70	26	96.91 ± 0.35	34	97.45 ± 0.81
pro	72	91.00 ± 3.94	71	93.00 ± 1.94	66	92.00 ± 2.22	59	97.00 ± 0.83
spe	188	85.38 ± 4.49	300	84.23 ± 0.96	324	82.69 ± 2.77	256	86.54 ± 6.08
tit	84	77.82 ± 2.06	79	77.82 ± 0.90	83	77.82 ± 1.52	83	77.86 ± 1.42
ttt	451	95.58 ± 0.35	508	79.58 ± 5.08	150	93.79 ± 0.61	391	95.68 ± 0.32
vot	78	96.98 ± 0.19	88	96.98 ± 0.45	49	96.51 ± 0.97	69	97.21 ± 1.17
wea	32	85.71 ± 3.35	40	78.57 ± 18.43	14	71.43 ± 6.71	31	71.43 ± 6.71

esquema de combinação de regras de todos os classificadores encontrados. O conjunto de validação é utilizado para treinar o classificador ETL que combina os votos dos classificadores.

Na Tabela 6.9, apresentamos uma comparação das abordagens em termos de médias de acurácias das abordagens propostas. Também, nesta tabela, os

Tabela 6.8: Resultados para a abordagem Comitê BAS com votação ETL.

Instância	#Cl.	Acurácia	
aud	282	98.18	± 0.50
bca	284	77.14	± 4.56
car	461	97.79	± 0.19
cmc	307	69.32	± 0.36
crx	254	86.67	± 0.08
ech	248	93.08	± 2.35
gla	14	99.52	± 1.26
hab	182	77.33	± 1.11
hco	276	84.17	± 1.77
hep	331	86.67	± 1.85
len	80	91.67	± 10.77
mk2	132	66.50	± 1.43
mk3	150	97.82	± 0.10
pro	317	96.00	± 2.29
spe	1410	86.92	± 0.43
tit	387	77.86	± 1.18
ttt	1382	96.32	± 1.03
vot	343	97.44	± 0.06
wea	133	78.57	± 5.03

melhores resultados para cada instância são apresentados em negrito.

Com o objetivo de demonstrar que a utilização de uma distribuição inicial não uniforme melhora o desempenho de classificadores gerados através de abordagens de Boosting, as abordagens propostas são comparadas com outros algoritmos de aprendizado de máquina relacionados. Os algoritmos comparados são enumerados a seguir.

Classificador-Base: o mesmo utilizado em cada abordagem Boosting.

AdaBoost: o algoritmo Adaboost original, que utiliza a mesma configuração de classificador-base.

Bagging de BAS: uma estratégia de comitê onde a determinação dos pesos iniciais dos exemplos é completamente aleatória.

Primeiramente, podemos perceber que, em todos os experimentos, as novas abordagens de Boosting propostas sempre produzem um classificador com melhor desempenho que o classificador-base, diferentemente das estratégias AdaBoost e Bagging de BAS, conforme ilustrado pelas instâncias *ech*, *hco*, *hep*, *len* e *mk2*.

Além disso, na grande maioria dos casos, todas as novas abordagens possuem melhores resultados que as estratégias convencionais. Tal fato não se observa somente em alguns poucos conjuntos. A seguir, destacamos alguns resultados específicos para certos conjuntos.

Tabela 6.9: Acurácias Médias para todas as abordagens supervisionadas.

Inst.	Classificador-Base	AdaBoost	Bagging de BAS	BAS Genético	Comitê BAS Melhor Maioria	Comitê BAS Votação ETL
aud	94.55 ± 3.48	95.45 ± 2.52	95.45 ± 1.26	97.27 ± 0.50	97.73 ± 0.63	98.18 ± 0.50
bca	71.07 ± 7.99	71.07 ± 2.08	71.43 ± 0.00	74.29 ± 0.20	73.57 ± 2.58	77.14 ± 4.56
car	92.27 ± 2.02	96.57 ± 0.15	96.22 ± 0.25	97.85 ± 0.05	97.21 ± 0.13	97.79 ± 0.19
cmc	63.61 ± 2.36	68.91 ± 0.51	67.35 ± 0.95	68.16 ± 0.15	68.71 ± 0.00	69.32 ± 0.36
crx	85.51 ± 4.76	86.09 ± 1.85	85.36 ± 0.76	86.23 ± 1.81	86.52 ± 0.12	86.67 ± 0.08
ech	91.54 ± 6.37	88.46 ± 1.07	90.00 ± 0.64	92.31 ± 2.13	93.08 ± 2.35	93.08 ± 2.35
gla	99.05 ± 1.67	99.05 ± 1.67	99.52 ± 1.26	99.52 ± 1.26	99.52 ± 1.26	99.52 ± 1.26
hab	71.00 ± 5.85	72.67 ± 1.67	73.00 ± 1.76	75.67 ± 0.65	77.00 ± 1.02	77.33 ± 1.11
hco	81.39 ± 5.57	80.28 ± 2.24	81.94 ± 1.16	83.06 ± 0.69	84.17 ± 1.00	84.17 ± 1.77
hep	82.00 ± 9.45	78.00 ± 2.41	80.00 ± 1.85	84.67 ± 1.30	84.00 ± 2.59	86.67 ± 1.85
len	87.50 ± 19.00	75.00 ± 6.34	79.17 ± 5.28	91.67 ± 2.12	91.67 ± 10.56	91.67 ± 10.77
mk2	65.67 ± 3.28	58.17 ± 0.88	63.67 ± 1.02	65.83 ± 2.55	65.83 ± 0.98	66.50 ± 1.43
mk3	75.09 ± 3.57	95.09 ± 0.35	96.55 ± 0.55	97.09 ± 0.70	97.45 ± 0.81	97.82 ± 0.10
pro	67.00 ± 7.90	88.00 ± 0.55	89.00 ± 0.28	94.00 ± 1.57	97.00 ± 0.83	96.00 ± 2.29
spe	80.00 ± 8.78	81.92 ± 2.46	83.85 ± 0.21	83.08 ± 1.49	86.54 ± 6.08	86.92 ± 0.43
tit	77.64 ± 2.05	77.86 ± 1.47	77.73 ± 0.12	77.82 ± 0.35	77.86 ± 1.42	77.86 ± 1.18
ttt	70.00 ± 4.38	94.42 ± 0.68	93.58 ± 0.32	95.79 ± 0.59	95.68 ± 0.32	96.32 ± 1.03
vot	95.58 ± 2.32	96.05 ± 0.19	96.98 ± 1.10	97.21 ± 0.52	97.21 ± 1.17	97.44 ± 0.06
wea	14.29 ± 30.71	71.43 ± 6.71	57.14 ± 13.40	78.57 ± 5.03	71.43 ± 6.71	78.57 ± 5.03

Conjunto cmc: onde os algoritmos BAS Genético e Comitê BAS com votação pela melhor maioria possuem pior desempenho que o AdaBoost. O algoritmo Comitê BAS com Votação ETL, em contrapartida, possui melhor desempenho.

Conjunto gla: onde todos as novas abordagens possuem desempenho equivalente ao algoritmo Bagging de BAS.

Conjunto spe: onde o algoritmo BAS Genético possui pior desempenho que o algoritmo Bagging de BAS.

Conjunto tit: onde o algoritmo BAS Genético possui pior desempenho que o algoritmo AdaBoost e as outras abordagens possuem desempenho equivalente.

Conjunto wea: onde o algoritmo Comitê BAS com votação pela melhor maioria possui desempenho equivalente ao algoritmo AdaBoost.

Por exemplo, no Conjunto *pro*, o ganho absoluto do Comitê BAS sobre o algoritmo AdaBoost é de 9%. Tal melhora em desempenho é considerável, visto que o AdaBoost também possui um considerável ganho absoluto em relação ao classificador-base de 21%.

Na grande maioria dos conjuntos, a abordagem Comitê BAS com votação ETL possui o melhor desempenho, com exceção do Conjunto *car*, onde a abordagem BAS Genético possui o melhor desempenho e no Conjunto *pro*, onde a abordagem Comitê BAS com votação pela melhor maioria foi a campeã.

Além disso, cabe ressaltar que a abordagem BAS Genético possui o menor desvio padrão. Na média, ele é por volta de metade do valor das outras abordagens.

Finalmente, na tabela 6.10, apresentamos uma comparação entre as abordagens experimentadas em termos de classificadores-base constituintes do comitê gerado. Tal comparação é importante nos casos em que desejamos gerar um classificador final rápido, com uma menor quantidade de membros básicos, mas que ainda possua um desempenho aceitável.

Tabela 6.10: Quantidade total de classificadores-base por abordagem.

Instância	Bagging		BAS	Comitê BAS	Comitê BAS
	AdaBoost	de BAS	Genético	Melhor Maioria	Votação ETL
aud	40	130	19	61	282
bca	188	101	16	40	284
car	80	478	80	199	461
cmc	135	424	70	152	307
crx	209	272	64	139	254
ech	50	145	24	56	248
gla	1	7	1	3	14
hab	169	105	10	29	182
hco	163	239	25	50	276
hep	74	206	33	57	331
len	107	88	11	18	80
mk2	21	64	9	20	132
mk3	60	162	18	34	150
pro	29	140	24	72	317
spe	380	250	26	188	1410
tit	28	85	10	84	387
ttt	149	1050	171	451	1382
vot	54	189	24	78	343
wea	43	115	10	32	133

6.5.2

Experimentos com Abordagens Semi-Supervisionadas

Agora apresentamos os experimentos com a abordagem semi-supervisionada.

O objetivo aqui é gerar classificadores melhores utilizando poucos dados rotulados e uma grande quantidade de amostras. A abordagem proposta é comparada com outros algoritmos de aprendizado de máquina relacionados. Os algoritmos comparados são enumerados a seguir.

Classificador-Base: o mesmo utilizado em cada abordagem Boosting.

AdaBoost: o algoritmo AdaBoost original, que utiliza a mesma configuração de classificador-base.

Comitê BAS Supervisionado: a mesma estratégia BAS, porém sem a utilização das amostras disponíveis.

Na Tabela 6.11, apresentamos uma comparação, em termos de acurácias médias e desvio padrão, para cada conjunto de treinamento e teste, destas abordagens. De novo, os melhores resultados para cada instância são apresentados em negrito.

Tabela 6.11: Acurácias médias para todas as abordagens comparadas.

Instância	Classificador-Base	AdaBoost	Comitê BAS Supervisionado	Comitê BAS Semi-Supervisionado
ban	84.44 ± 1.47	84.44 ± 1.47	85.80 ± 0.92	87.07 ± 0.54
bca	70.34 ± 3.52	73.42 ± 3.89	73.71 ± 4.01	76.96 ± 3.83
dia	71.90 ± 1.82	75.05 ± 1.42	75.75 ± 1.64	77.34 ± 1.19
fls	56.12 ± 1.82	67.61 ± 1.54	67.73 ± 1.62	67.74 ± 1.61
ger	70.18 ± 1.72	75.31 ± 2.12	76.59 ± 1.22	76.49 ± 1.60
hea	72.96 ± 2.99	80.55 ± 3.13	82.25 ± 3.16	85.15 ± 2.69
ima	87.67 ± 1.68	92.49 ± 2.31	95.20 ± 1.20	96.33 ± 0.45
rin	61.23 ± 0.97	91.82 ± 0.70	92.41 ± 0.55	92.17 ± 0.60
spl	77.00 ± 1.39	93.32 ± 0.40	93.62 ± 0.36	93.73 ± 0.33
tit	77.32 ± 2.82	76.75 ± 2.66	77.32 ± 2.38	77.86 ± 1.73
thy	78.00 ± 1.25	93.89 ± 1.28	94.23 ± 1.17	96.21 ± 0.56
two	66.47 ± 1.19	92.46 ± 0.56	93.49 ± 0.59	94.16 ± 0.38
wav	74.54 ± 2.37	86.72 ± 0.56	86.90 ± 0.59	87.36 ± 0.38

Podemos perceber pela Tabela 6.11 que as abordagens Boosting baseadas no algoritmo BAS mantém a característica de possuir melhor desempenho que o algoritmo AdaBoost original.

Em alguns casos esse aumento de desempenho é notável. Por exemplo, na instância *ima*, o ganho absoluto é de aproximadamente 2.7%. Tal ganho é considerável visto que o algoritmo AdaBoost já possui uma acurácia relativamente alta, aproximadamente 92.5. Além disso, a inclusão de amostras nesta tarefa proporciona um desempenho ainda maior, onde o ganho relativo é de aproximadamente 3.8%

Na grande maioria das tarefas, a inclusão de amostras gerou classificadores com melhor acurácia. Apenas em duas tarefas, *ger* e *rin*, o desempenho foi pior, com uma perda absoluta da ordem de 0.1% e 0.23%, respectivamente.

Na Tabela 6.12, é conduzida uma comparação da abordagem proposta com outras abordagens supervisionadas de aprendizado de máquina reportadas (Mik99) para as mesma tarefas, que utilizam a mesma estratégia de experimentação para cada conjunto de treinamento e teste.

Como pode ser visto na Tabela 6.12, a abordagem possui um desempenho competitivo em relação ao melhor classificador reportado para cada tarefa. Tal fato é um excelente resultado pois não estamos comparando a estratégia proposta com apenas uma solução baseada em aprendizado de máquina, porém, com a melhor entre três possibilidades, AdaBoost Regularizado (ABR),

Tabela 6.12: Comparação da abordagem semi-supervisionada com alguns algoritmos de estado-da-arte.

Instância	Melhor Resultado (Mik99)		Comitê BAS Semi-Supervisionado
	ban	KFD	89.20 \pm 0.50
bca	KFD	74.20 \pm 4.60	76.96 \pm 3.83
dia	KFD	76.80 \pm 1.60	77.34 \pm 1.19
fls	SVM	67.60 \pm 1.80	67.74 \pm 1.61
ger	SVM	76.40 \pm 2.10	76.49 \pm 1.60
hea	SVM	84.00 \pm 3.30	85.15 \pm 2.69
ima	ABR	97.30 \pm 0.60	96.33 \pm 0.45
rin	KFD	98.50 \pm 0.10	92.17 \pm 0.60
spl	ABR	90.50 \pm 0.70	93.73 \pm 0.33
tit	SVM	76.40 \pm 2.10	77.86 \pm 1.73
thy	KFD	95.80 \pm 1.00	96.21 \pm 0.56
two	KFD	97.40 \pm 0.20	94.16 \pm 0.38
wav	ABR	90.20 \pm 0.80	87.36 \pm 0.38

Máquinas de Vetores de Suporte (SVM) e Discriminante de Núcleo de Fischer (KFD) (Mik99).

Finalmente, na Figura 6.1, é apresentada uma comparação entre as diferentes abordagens do Comitê BAS para a instância *bca*, no tocante à quantidade de exemplos e amostras para a abordagem supervisionada.

No gráfico da Figura 6.1, o eixo Y representa a acurácia média nos conjuntos de teste e o eixo X representa a quantidade de exemplos do conjunto de treinamento em centenas de exemplos. Como a quantidade total de amostras, neste caso, não é infinita, a quantidade de amostras diminui proporcionalmente com o aumento do número de exemplos, até atingir o valor zero. Para tal quantidade de amostras, as abordagens se tornam equivalentes. Podemos perceber, pelo gráfico, que para obter, com a abordagem supervisionada, o mesmo desempenho que a abordagem semi-supervisionada inicial, é necessário um conjunto de exemplos sessenta vezes maior (200 / 12000). Tal fato demonstra a importância do auxílio do conjunto de amostras no aprendizado do Comitê BAS para conjuntos de dados pequenos. Comportamentos semelhantes são observados também nas outras instâncias analisadas, onde o desempenho da abordagem semi-supervisionada é superior.

6.6 Sumário

Este capítulo apresenta alguns resultados empíricos da aplicação das diferentes estratégias BAS propostas, na solução de tarefas de classificação de dados. Dois conjuntos de dados foram utilizados, o Conjunto do Repositório

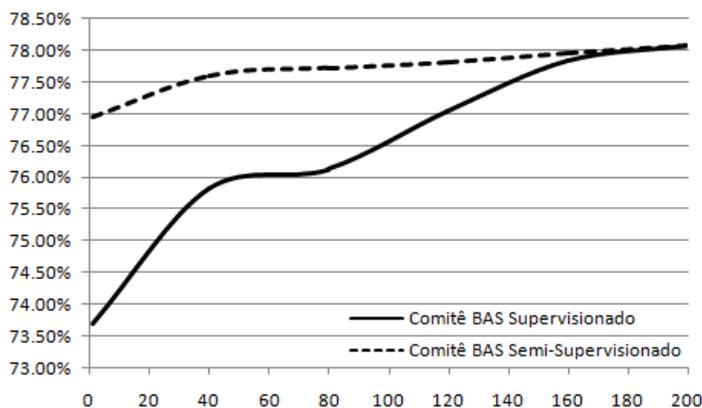


Figura 6.1: Comparação entre as estratégias supervisionada e semi-supervisionada para a instância *bca*.

UCI e o Conjunto do Repositório IDA, para, respectivamente, avaliar as abordagens supervisionadas e semi-supervisionadas.

Foram apresentados resultados de acurácias médias e desvio padrão em termos de iterações de validação cruzada realizadas. Resultados não muito diferentes são obtidos caso a validação cruzada seja aplicada diversas vezes para o mesmo conjunto. Tal fato se observa pois o caráter aleatório das abordagens BAS é responsável por uma pequena parte do algoritmo e, além disso, é compensado pela utilização de um conjunto de validação à parte.

Utilizando uma configuração padrão para os parâmetros, as abordagens propostas tiveram sistematicamente um desempenho melhor que o algoritmo AdaBoost. Em alguns casos mostramos também que os resultados, mesmo utilizando uma modelagem extremamente simples, são competitivos em termos de desempenho com outros algoritmos de aprendizado de máquina.

Em termos de desempenho computacional, as abordagens também são eficientes. A abordagem Comitê BAS pela Melhor Maioria processa 42.000 exemplos por segundo em média. Esta velocidade cai para 39.000 exemplos por segundo quando é utilizado o esquema de Votação ETL.

Já no caso da abordagem BAS Genético, essa velocidade é de apenas 2.000 exemplos por segundo. Entretanto, tal degradação, em termos de tempo de treinamento e desempenho no conjunto de teste, é compensada na extração, onde apenas um classificador BAS é empregado.

Os resultados aqui apresentados indicam que a utilização de uma distribuição não uniforme em uma abordagem BAS é um método efetivo e econômico de gerar classificadores para tarefas de classificação de dados.