

Diversidade

Roberto Pinheiro

Roteiro

- Motivação
- Introdução
- Medidas de Diversidade
 - Pareadas
 - Não Pareadas
 - Experimentos
- Métodos de Criação de Diversidade
- Conclusões
- Referências

Motivação

- Não existe classificador perfeito
- Classificadores **diversos** acertam e erram em situações diferentes
- Diversidade parece ser um **requerimento** para o sucesso de uma combinação

Introdução

- Apesar de ser um conceito aparentemente intuitivo
- Não existe uma definição formalmente aceita de diversidade
- Isso não impediu pesquisadores de insistirem em alcançar a diversidade

Medidas de Diversidade

- Pareadas
 - Envolvem apenas dois classificadores
- Não Pareadas
 - Envolvem três ou mais classificadores

Medidas de Diversidade

Pareadas

- Suponha dois Classificadores α e β
- Tabela de Relação entre os Classificadores

	Classificador α acerta	Classificador α erra
Classificador β acerta	a	b
Classificador β erra	c	d

- a, b, c e d são valores inteiros representando a quantidade de padrões de teste nas condições especificadas

Q Statistics [5]

$$Q_{\alpha,\beta} = \frac{ad - bc}{ad + bc}$$

- $Q_{\alpha,\beta}$ varia entre -1 e 1
 - Reconhecem os mesmos padrões tende à +1
 - Cometem erros diferentes tendem à -1
- $Q_{\alpha,\beta} = 0$, para classificadores estatisticamente independentes

Correlation Coefficient ρ [6]

$$\rho_{\alpha,\beta} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

- $Q_{\alpha,\beta}$ e $\rho_{\alpha,\beta}$ terão o mesmo sinal
- $|Q_{\alpha,\beta}| \geq |\rho_{\alpha,\beta}|$

Disagreement Measure [7]

$$Dis_{\alpha, \beta} = \frac{b + c}{a + b + c + d}$$

- Probabilidade dos classificadores α e β discordarem da decisão
- *Agreement Measure* é o inverso desta

Double-Fault Measure [8]

$$DF_{\alpha, \beta} = \frac{d}{a + b + c + d}$$

- Probabilidade dos classificadores α e β errarem juntos

Medidas de Diversidade

Não Pareadas

- Calcula a Diversidade entre 3 ou mais classificadores
- É possível utilizar as medidas pareadas para obter a diversidade de um conjunto de classificadores (valor médio)

$$Medida_{av} = \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^L Medida_{i,j}$$

Entropy Measure E [9]

$$E = \frac{1}{N} \frac{2}{L-1} \sum_{j=1}^N \min \left\{ \underbrace{\left(\sum_{i=1}^L y_{j,i} \right)}_{\substack{\text{Quantidade de votos} \\ \text{corretos para} \\ \text{um padrão}}}, \underbrace{\left(L - \sum_{i=1}^L y_{j,i} \right)}_{\substack{\text{Quantidade de votos} \\ \text{incorretos para} \\ \text{um padrão}}} \right\}$$

0 -> nenhuma diferença
1 -> maior diversidade possível

- **Intuitivamente** um *ensemble* é mais diverso quando os $L/2$ classificadores acertam e os outros $L/2$ dos classificadores erram
 - Afinal, se todos acertarem tudo ou errarem tudo, não há desacordo. Logo, não vale a pena combiná-los.
- Apesar do nome, não é a Entropia usual

Entropy Measure E [9]

$$E = \frac{1}{N} \frac{2}{L-1} \sum_{j=1}^N \min \left\{ \underbrace{\left(\sum_{i=1}^L y_{j,i} \right)}_{\text{Quantidade de votos corretos para um padrão}}, \underbrace{\left(L - \sum_{i=1}^L y_{j,i} \right)}_{\text{Quantidade de votos incorretos para um padrão}} \right\}$$

0 -> nenhuma diferença

1 -> maior diversidade possível

corretos para
um padrão

incorretos para
um padrão

- $y_{j,i} = 1$, se o classificador i acertou o padrão j
- $y_{j,i} = 0$, caso contrário
- $L-1$ na divisão parece incorreto
 - Exemplo $N = 3$ e $L = 4$

Kohavi-Wolpert Variance [10]

$$KW = \frac{1}{NL^2} \sum_{j=1}^N \left(\sum_{i=1}^L y_{j,i} \right) \left(L - \sum_{i=1}^L y_{j,i} \right)$$

Quantidade de votos

corretos para
um padrão

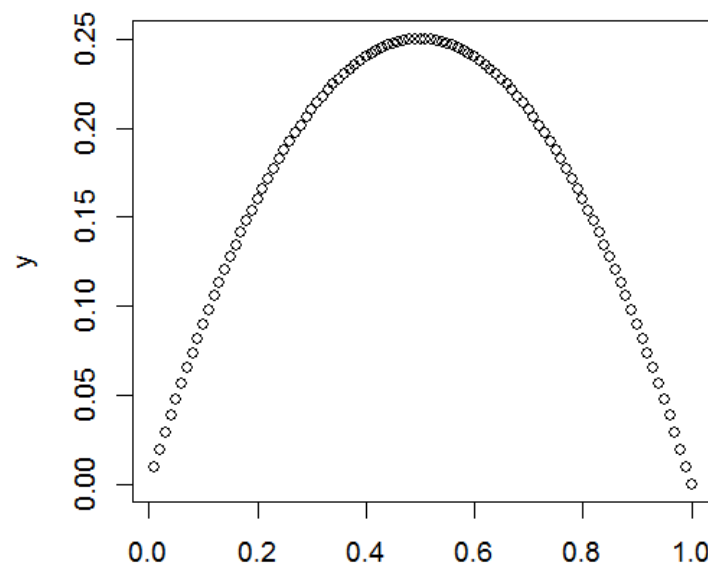
Quantidade de votos

incorretos para
um padrão

0.00 -> nenhuma diferença

0.25 -> maior diversidade possível

- Mesma ideia que a Entropia
 - Mais diversidade quando os classificadores dividem os acertos e erros



Measure of Interrater Agreement [11]

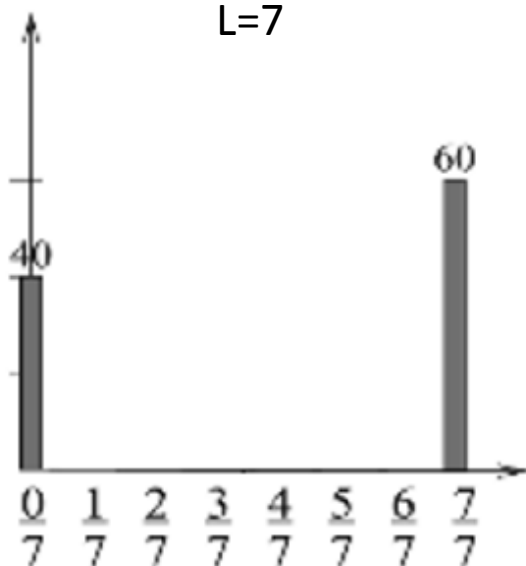
$$\kappa = 1 - \frac{\frac{1}{L} \sum_{j=1}^N \left(\sum_{i=1}^L y_{j,i} \right) \left(L - \sum_{i=1}^L y_{j,i} \right)}{N(L-1)\bar{p}(1-\bar{p})}$$

- No qual: $\bar{p} = \frac{1}{NL} \sum_{j=1}^N \sum_{i=1}^L y_{j,i}$
 - 1 -> nenhuma diferença
 - 1 -> maior diversidade possível
 - Média da taxa de acerto dos classificadores do *ensemble*
- Existe uma relação com outras medidas

$$\kappa = 1 - \frac{L}{(L-1)\bar{p}(1-\bar{p})} KW = 1 - \frac{1}{2\bar{p}(1-\bar{p})} D_{av}$$

Measure of “difficulty” θ [12]

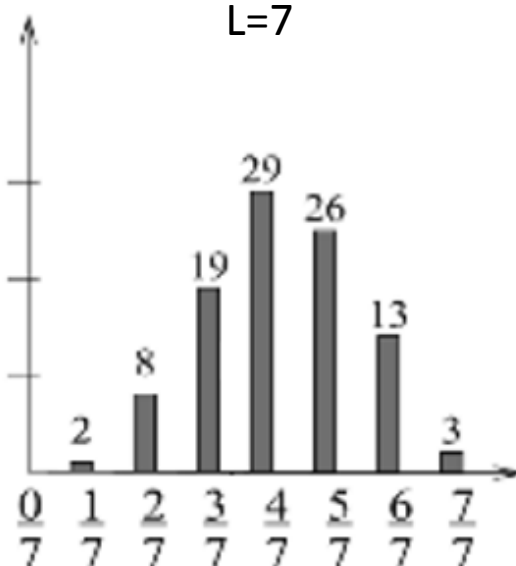
L=7



Variância $\rightarrow 0.240$

Baixa diversidade

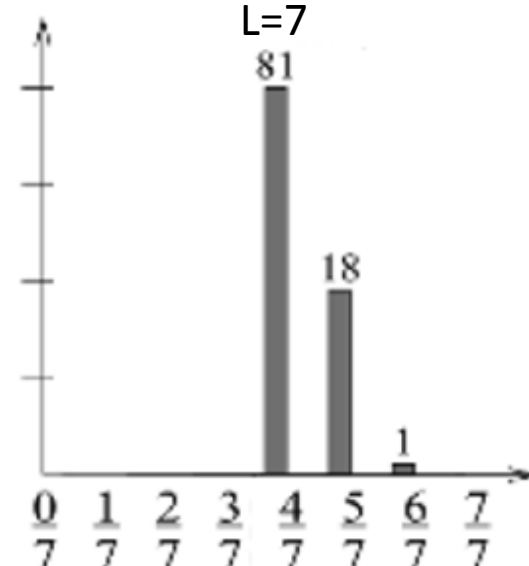
L=7



Variância $\rightarrow 0.034$

Diversidade intermediária

L=7



Variância $\rightarrow 0.004$

Maior diversidade

- Se a variância e média forem pequenas?
 - Exemplo: 40 em 0/7 e 60 em 1/7 = 0.004752

Generalized Diversity [13]

$$GD = 1 - \frac{\sum_{i=1}^L \frac{i(i-1)}{L(L-1)} p_i}{\sum_{i=1}^L \frac{i}{L} p_i}$$

p_i é a probabilidade de
exatamente i
classificadores
aleatórios **errarem**

- Supondo apenas 2 classificadores
 - Diversidade é **máxima** (1) quando a **falha** de um dos classificadores é sempre acompanhada pelo **acerto** do outro classificador
 - Diversidade é **mínima** (0) quando a **falha** de um dos classificadores é sempre acompanhada pela **falha** do outro classificador

Coincident Failure Diversity [13]

$$CFD = \begin{cases} 0, & p_0 = 1 \\ \frac{1}{1 - p_0} \sum_{i=1}^L \frac{L-i}{L-1} p_i, & p_0 < 1 \end{cases}$$

- Modificação do *GD*
- Insere probabilidade de nenhum classificador errar (p_0)
 - Diversidade é **mínima** (0) se todos classificadores, simultaneamente, sempre acertam ou erram
 - Diversidade é **máxima** (1) se todo erro é cometido por no máximo um classificador

Taxa de Acerto X Diversidade

- É **esperado** que quanto maior a diversidade entre os classificadores de uma combinação, melhor será sua taxa de acerto
- Essa expectativa é a realidade?

Taxa de Acerto X Diversidade

- Um estudo foi realizado sobre uma combinação de três classificadores [3]
- Foram utilizados
 - 10 medidas de diversidade diferentes
 - Medidas pareadas foram condensadas pela média
 - 10 padrões de teste
 - 28 combinações de classificadores diferentes
 - Cada classificador individual com taxa de acerto igual 0.6
- Resultado final obtido por voto majoritário

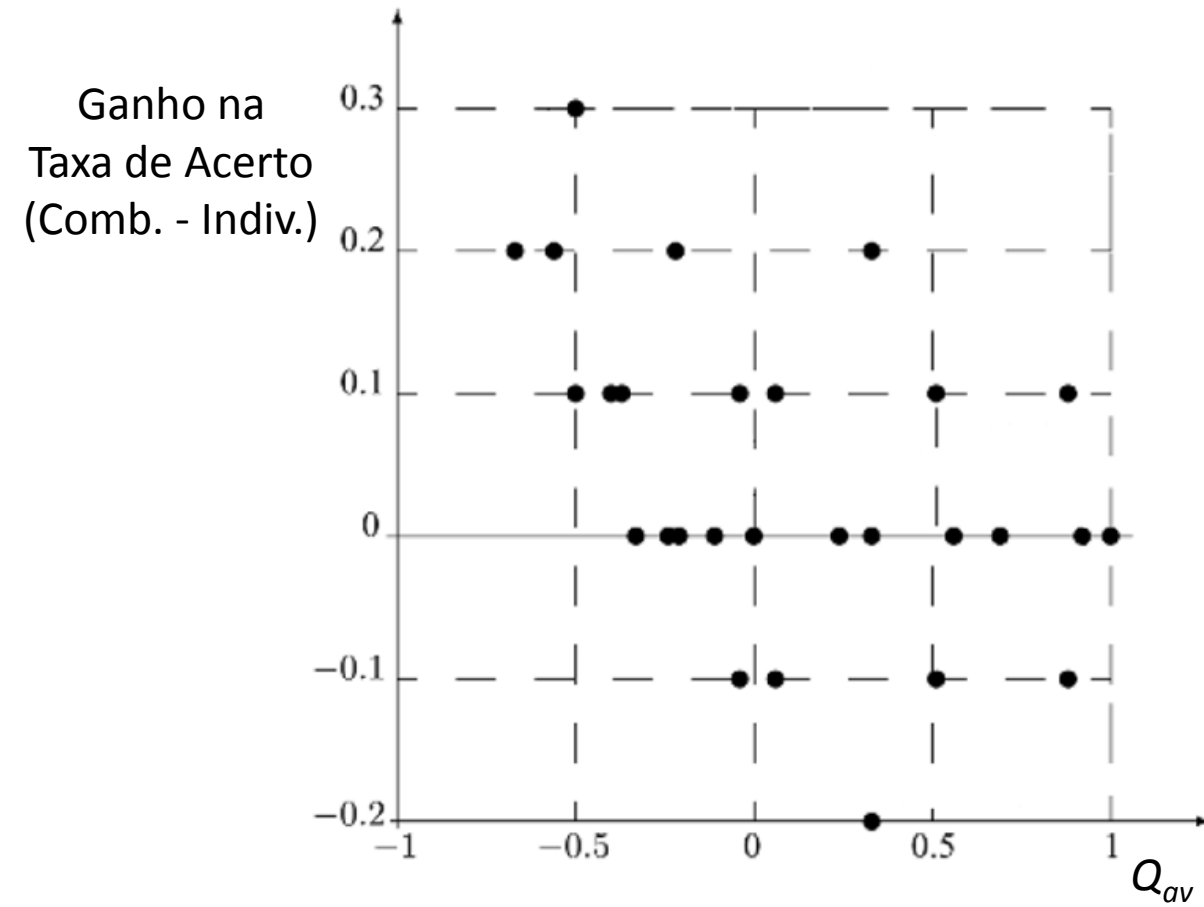
No.	<i>Q</i>	<i>ρ</i>	<i>D</i>	<i>DF</i>	<i>KW</i>	<i>κ</i>	<i>E</i>	<i>θ</i>	<i>GD</i>	<i>CFD</i>	Voto
1	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.90	0.9
2	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.75	0.8
3	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.83	0.8
4	-0.67	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
5	-0.56	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
6	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.50	0.7
7	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.64	0.7
8	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
9	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
10	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
11	-0.39	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
12	-0.38	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
13	1.0	1.0	0.00	0.40	0.00	1.0	0.00	0.24	0.00	0.00	0.6
14	0.92	0.72	0.13	0.33	0.04	0.72	0.20	0.20	0.17	0.30	0.6
15	0.69	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
16	0.56	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
17	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
18	0.24	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
19	0.00	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
20	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
21	-0.11	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
22	-0.21	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
23	-0.33	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
24	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.30	0.5
25	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.50	0.5
26	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
27	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
28	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.50	0.4

No.	Q	ρ	D	DF	KW	κ	E	θ	GD	CFD	Voto
1	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.90	0.9
2	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.75	0.8
3	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.83	0.8
4	-0.67	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
5	-0.56	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
6	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.50	0.7
7	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.64	0.7
8	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
9	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
10	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
11	-0.39	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
12	-0.38	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
13	1.0	1.0	0.00	0.40	0.00	1.0	0.00	0.24	0.00	0.00	0.6
14	0.92	0.72	0.13	0.33	0.04	0.72	0.20	0.20	0.17	0.30	0.6
15	0.69	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
16	0.56	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
17	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
18	0.24	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
19	0.00	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
20	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
21	-0.11	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
22	-0.21	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
23	-0.33	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
24	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.30	0.5
25	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.50	0.5
26	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
27	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
28	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.50	0.4

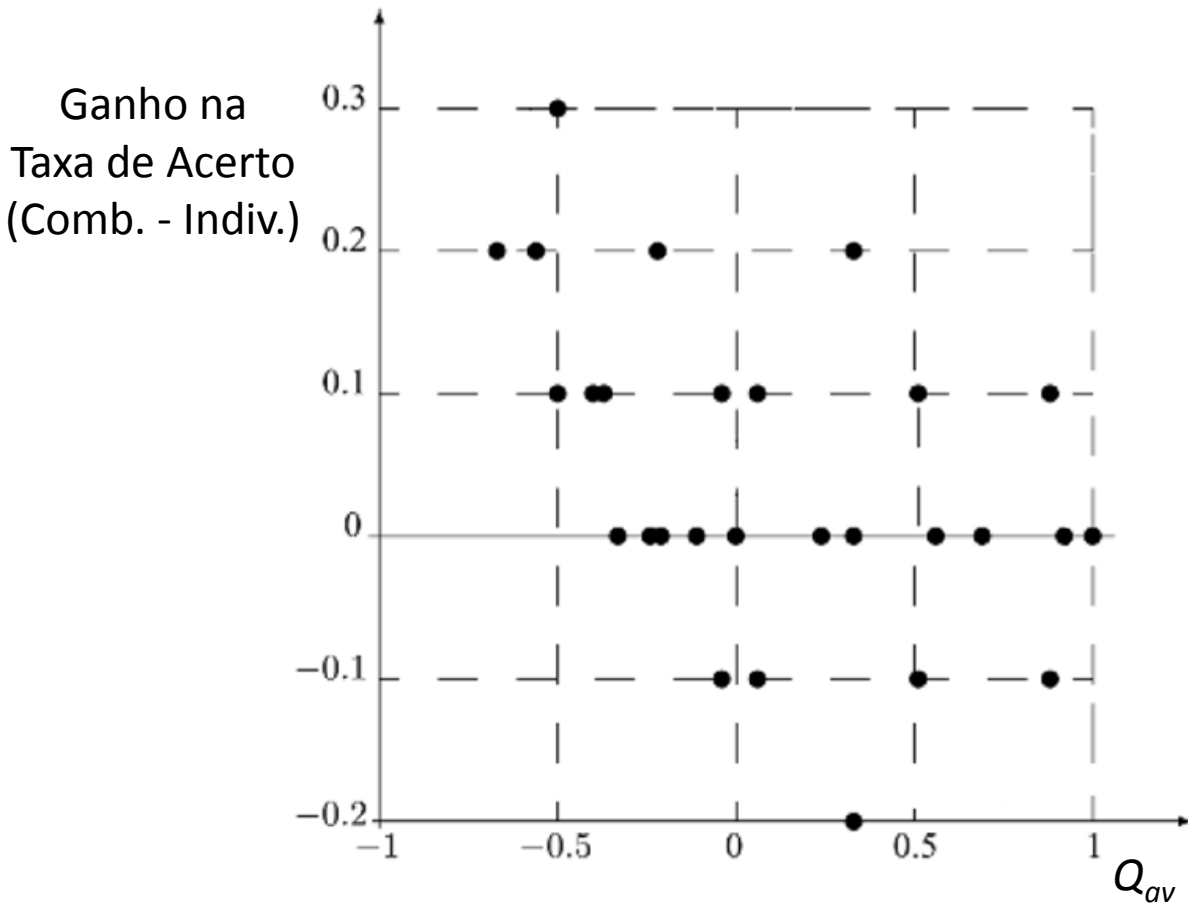
No.	<i>Q</i>	<i>ρ</i>	<i>D</i>	<i>DF</i>	<i>KW</i>	<i>κ</i>	<i>E</i>	<i>θ</i>	<i>GD</i>	<i>CFD</i>	Voto
1	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.90	0.9
2	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.75	0.8
3	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.83	0.8
4	-0.67	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
5	-0.56	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
6	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.50	0.7
7	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.64	0.7
8	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
9	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
10	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
11	-0.39	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
12	-0.38	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
13	1.0	1.0	0.00	0.40	0.00	1.0	0.00	0.24	0.00	0.00	0.6
14	0.92	0.72	0.13	0.33	0.04	0.72	0.20	0.20	0.17	0.30	0.6
15	0.69	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
16	0.56	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
17	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
18	0.24	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
19	0.00	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
20	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
21	-0.11	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
22	-0.21	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
23	-0.33	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
24	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.30	0.5
25	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.50	0.5
26	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
27	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
28	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.50	0.4

No.	<i>Q</i>	<i>ρ</i>	<i>D</i>	<i>DF</i>	<i>KW</i>	<i>κ</i>	<i>E</i>	<i>θ</i>	<i>GD</i>	<i>CFD</i>	Voto
1	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.90	0.9
2	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.75	0.8
3	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.83	0.8
4	-0.67	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
5	-0.56	-0.39	0.67	0.07	0.22	-0.39	1.0	0.02	0.83	0.90	0.8
6	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.50	0.7
7	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.64	0.7
8	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
9	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.75	0.7
10	-0.50	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
11	-0.39	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
12	-0.38	-0.25	0.60	0.10	0.20	-0.25	0.90	0.04	0.75	0.83	0.7
13	1.0	1.0	0.00	0.40	0.00	1.0	0.00	0.24	0.00	0.00	0.6
14	0.92	0.72	0.13	0.33	0.04	0.72	0.20	0.20	0.17	0.30	0.6
15	0.69	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
16	0.56	0.44	0.27	0.27	0.09	0.44	0.40	0.15	0.33	0.50	0.6
17	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
18	0.24	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
19	0.00	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.64	0.6
20	-0.22	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
21	-0.11	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
22	-0.21	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
23	-0.33	-0.11	0.53	0.13	0.18	-0.11	0.80	0.06	0.67	0.75	0.6
24	0.88	0.58	0.20	0.30	0.07	0.58	0.30	0.17	0.25	0.30	0.5
25	0.51	0.31	0.33	0.23	0.11	0.31	0.50	0.13	0.42	0.50	0.5
26	0.06	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
27	-0.04	0.03	0.47	0.17	0.16	0.03	0.70	0.08	0.58	0.64	0.5
28	0.33	0.17	0.40	0.20	0.13	0.17	0.60	0.11	0.50	0.50	0.4

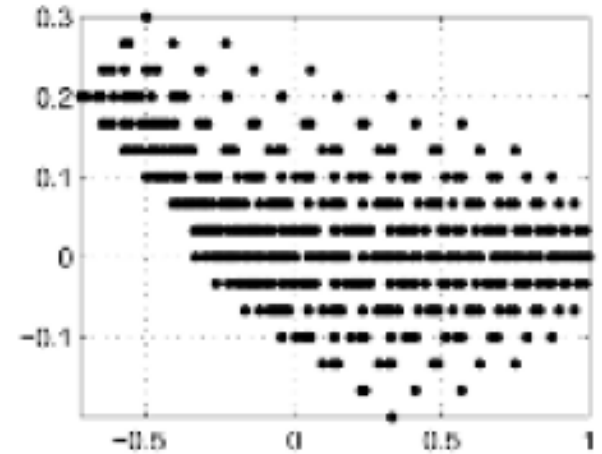
Taxa de Acerto X Diversidade



Taxa de Acerto X Diversidade

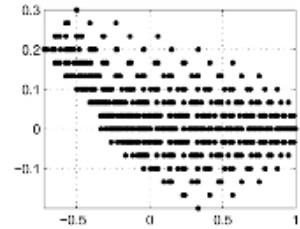


- Aumentar a quantidade de combinações facilita a visualização de uma “relação”:

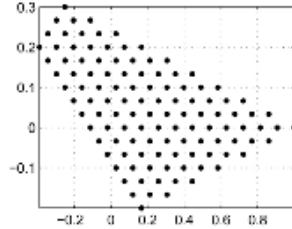


Taxa de Acerto X Diversidade

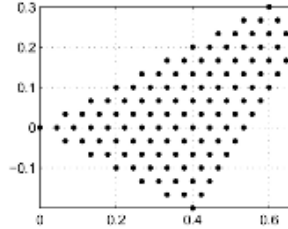
Q



ρ



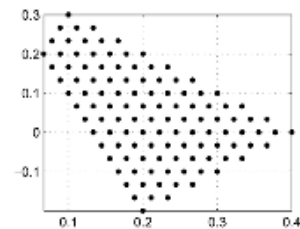
D



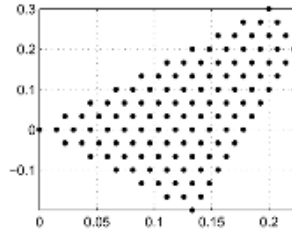
- 563 combinações de classificadores

- Comportamento similar entre as medidas de similaridade

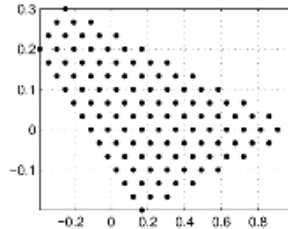
DF



KW

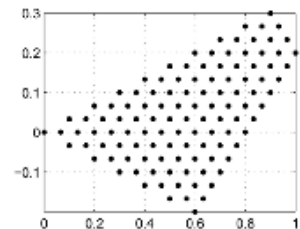


κ

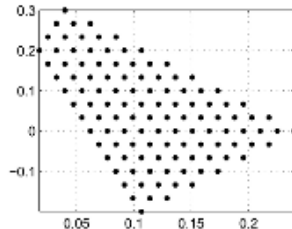


- Uma medida isolada não é precisa o suficiente para capturar a diversidade
 - Talvez medidas combinadas sejam uma boa solução

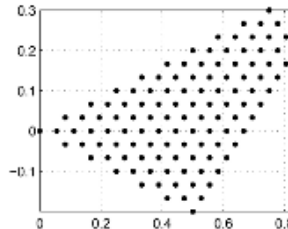
E



θ



GD



- **Existe um limiar para cada medida que garante uma melhoria**

Correlação entre as medidas

	ρ	Dis	DF	KW	κ	E	θ	GD	CFD
Q	98	-97	65	-97	98	-97	36	-98	-95
ρ		-97	72	-97	100	-97	44	-100	-97
Dis			-56	100	-98	100	-26	98	96
DF				-56	70	-56	93	-72	-70
KW					-98	100	-26	98	96
κ						-98	41	-100	-98
E							-26	98	96
θ								-44	-42
GD									98

Correlação (em %) das medidas de diversidade em experimentos realizados sobre base *breast cancer*

Correlação entre as medidas

	ρ	Dis	DF	KW	κ	E	θ	GD	CFD
Q	98	-97	65	-97	98	-97	36	-98	-95
ρ		-97	72	-97	100	-97	44	-100	-97
Dis			-56	100	-98	100	-26	98	96
DF				-56	70	-56	93	-72	-70
KW					-98	100	-26	98	96
κ						-98	41	-100	-98
E							-26	98	96
θ								-44	-42
GD									98

Correlação (em %) das medidas de diversidade em experimentos realizados sobre base *breast cancer*

Considerações

- Precisamos de diversidade
 - A ideia de que é necessário classificadores diversos é bem clara e intuitiva
- Mas não sabemos calcular bem
 - Não há medida universal nem definição formal
 - As medidas atuais
 - Não exibem uma relação forte com a taxa de acerto da combinação
 - Na maioria dos casos se comportam similarmente
- Então, se as medidas não são confiáveis como obter Diversidade?

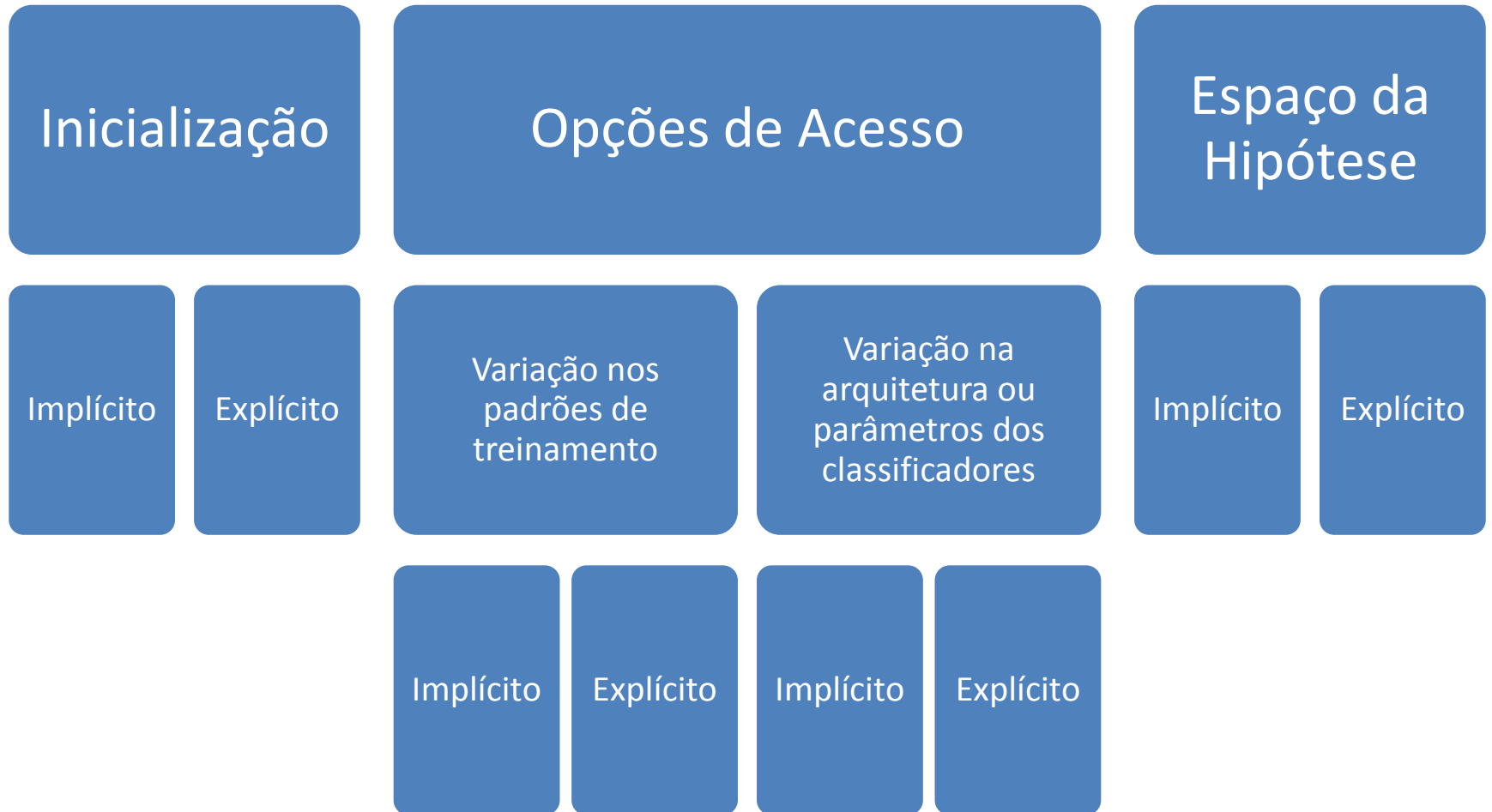
Métodos para Criação de Diversidade

- Existem vários métodos
- Usaremos a taxonomia criada por Brown [4] como guia
- Primeira distinção
 - Implícitos
 - Usa **aleatoriedade** na expectativa de gerar Diversidade
 - Explícitos
 - Utiliza **determinismo** para reforçar a presença da Diversidade
- Apesar de todos os métodos buscarem a Diversidade, dado a falta de formalização, nenhum deles **garante** Diversidade

Métodos para Criação de Diversidade

- Segunda distinção
 - Inicialização
 - Variação no ponto inicial influenciando a convergência
 - Opções de Acesso
 - Variação no que é utilizado para treinamento dos classificadores
 - Espaço da Hipótese
 - Variação na resposta do *ensemble* final

Taxonomia



Taxonomia – Inicialização

- Inicialização aleatória dos pesos para cada Rede Neural
 - Mais comum
 - Menos efetivo em gerar Diversidade
 - Convergência bastante similar
 - Implícito

Taxonomia – Inicialização

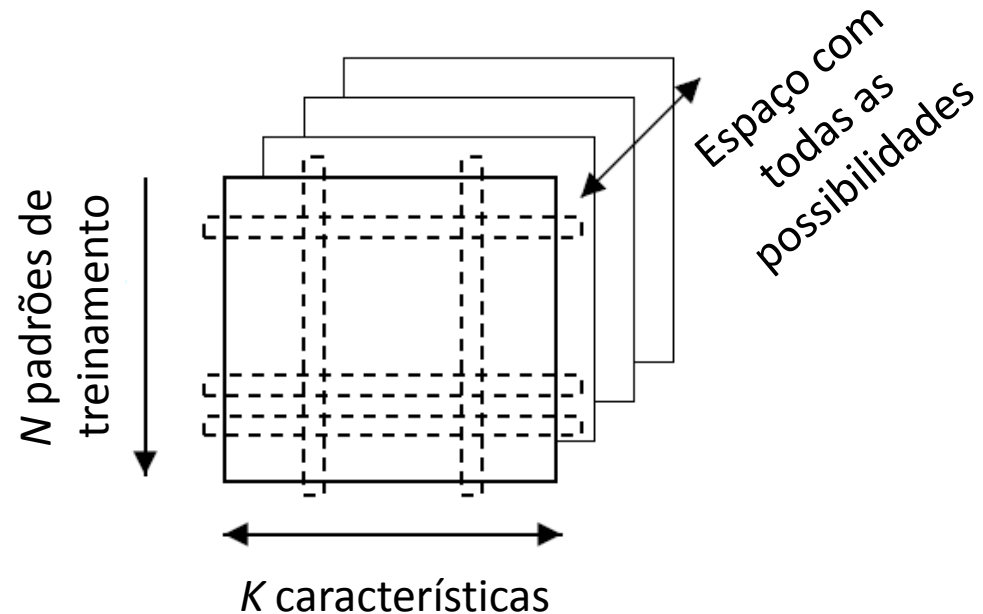
- Determinar pesos de modo que cada Rede Neural ocupe um lugar diferente no espaço de hipóteses
 - Levemente melhor que a versão aleatória
 - Explícito

Taxonomia – Inicialização

- *Fast Comitte Learning*
 - Usa apenas uma Rede Neural mas guarda os pesos de M diferentes estados e usa cada estado como uma Rede Neural do ensemble
 - Mais rápido do que treinar diferentes Redes Neurais
 - Implícito se usar estados aleatórios
 - Explícito se usar alguma métrica para definir os estados

Taxonomia – Opções de Acesso

- Variação nos padrões de treinamento
 - *Bagging*
 - *Random Subspace*
- Variação na arquitetura/parâmetros dos classificadores
 - Número de nós na camada escondida
 - Número de vizinhos usados para classificação



Taxonomia – Opções de Acesso

Variação nos padrões de treinamento

- Pré-processamento
 - Representação
 - Transformação
 - Normalização
 - Inserção de Ruído
 - Implícito / Explícito dependendo de como o procedimento foi realizado
- Modificar saída de alguns padrões de treinamento
 - Inserção de ruído aleatório
 - Segundo criador é mais efetivo que *Bagging*
 - Implícito

Taxonomia – Opções de Acesso

Variação nos padrões de treinamento

- AdaBoost
 - Altera a distribuição dos padrões de treinamento de cada classificador baseado nos erros do classificador anterior
 - É esperado que os classificadores “futuros” acertem onde os anteriores erraram e vice-versa
 - Explícito

Taxonomia – Opções de Acesso

Variação nos padrões de treinamento

- Medida de Diversidade
 - Usar alguma medida para encontrar subconjunto de padrões ou características que beneficiem a Diversidade
 - Construção **gradativa** do *ensemble* com inserção dos classificadores de acordo com a Diversidade atual
 - Só insere classificadores que aumentem Diversidade
 - Explícito

Taxonomia – Opções de Acesso

Variação nos padrões de treinamento

- *DECORATE*
 - Similar a Medida de Similaridade
 - Entretanto, trabalha com inserção de **padrões artificiais** em vez de modificação do conjunto de treinamento
 - Geram-se conjuntos discordantes de padrões artificiais que aumentem a taxa de acerto do ensemble
 - O processo é iterativo e termina após uma quantidade de iterações ou quando o tamanho desejado do *ensemble* tenha sido atingido
 - Explícito

Taxonomia – Opções de Acesso

Variação nos padrões de treinamento

- *Input Decimation Ensembles*
 - Usa seleção de características para que cada classificador seja especializado numa classe ou conjunto de classes
 - Mostrou grandes benefícios em dados reais e artificiais
 - Resultados melhores que *Bagging* e *Random Subspace*
 - Explícito

Taxonomia – Opções de Acesso

Variação na arquitetura/parâmetros dos classificadores

- Variação do número de nós
 - Pouco impacto nos experimentos realizados
 - Apenas uma base
 - Apenas variando entre 8-12 nós
 - Não gerou tanta Diversidade
 - Implícito se aleatório
 - Explícito se houver alguma metodologia

Espaço da Hipótese

- Métodos de Penalidade
 - Utilizar o erro para gerar *overfitting* em classificadores individuais
- Métodos Evolucionários
 - Utiliza algoritmos para buscar a melhor combinação
 - Função de *Fitness*: alguma Medida de Diversidade
 - Em vez de gerar indivíduos que explorem o máximo possível gera indivíduos que se complementam

Conclusão

- Diversidade ainda é um conceito informal
- Não há medida de diversidade universal
 - Além de não serem fortemente relacionadas com taxa de acerto
- Taxonomia abstrata e categorias abrangentes
 - Algumas categorias ainda estão pouco exploradas
- Existem muitas maneiras de obter diversidade
 - Métodos explícitos, em sua maioria, são superiores aos implícitos

Referências

- [1] Kuncheva, Ludmila I., and Christopher J. Whitaker. "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy." *Machine learning* 51.2 (2003): 181-207.
- [2] Zhou, Zhi-Hua. *Ensemble methods: foundations and algorithms*. CRC Press, 2012.
- [3] Kuncheva, Ludmila I. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.
- [4] Brown, Gavin, et al. "Diversity creation methods: a survey and categorisation." *Information Fusion* 6.1 (2005): 5-20.
- [5] Yule, G. Udny. "On the association of attributes in statistics: with illustrations from the material of the childhood society, &c." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* (1900): 257-319.

Referências

- [6] Sneath, Peter HA, and Robert R. Sokal. *Numerical taxonomy. The principles and practice of numerical classification*. 1973.
- [7] D. B. Skalak. The sources of increased accuracy for two proposed boosting algorithms. In *Proc. American Association for Artificial Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop*, 1996.
- [8] Ruta, Dymitr, and Bogdan Gabrys. "Analysis of the correlation between majority voting error and the diversity measures in multiple classifier systems." In *Proc. SOCO 2001* (2001): 50.
- [9] Cunningham, Padraig, and John Carney. "Diversity versus quality in classification ensembles based on feature selection." *Machine Learning: ECML 2000*. Springer Berlin Heidelberg, 2000. 109-116.

Referências

- [10] Kohavi, Ron, and David H. Wolpert. "Bias plus variance decomposition for zero-one loss functions." *ICML*. 1996.
- [11] Fleiss, Joseph L., Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.
- [12] Hansen, Lars Kai, and Peter Salamon. "Neural network ensembles." *IEEE transactions on pattern analysis and machine intelligence* 12.10 (1990): 993-1001.
- [13] Partridge, Derek, and W. Krzanowski. "Software diversity: practical statistics for its measurement and exploitation." *Information and software technology* 39.10 (1997): 707-717.