

Seleção Dinâmica de Classificadores

Thiago José Marques Moura

Centro de Informática
Universidade Federal de Pernambuco
www.cin.ufpe.br/~viisar

Setembro, 2015

Sumário

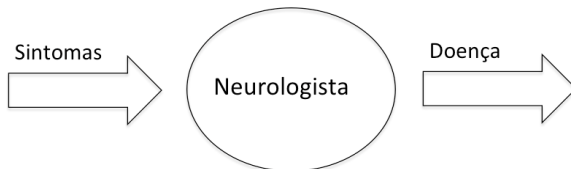
- 1 Introdução
- 2 Seleção Dinâmica
- 3 Métodos Individuais de DCS
- 4 Métodos de Grupo de DCS
- 5 Experimentos

Classificador Único - Ex. Médico

- Único Médico Generalista



- Único Médico Especialista



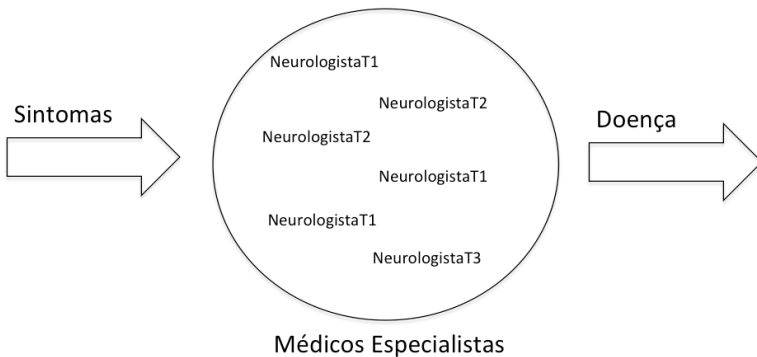
Ensemble de Classificadores - Ex. Médicos

- Comitê de Médicos



Ensemble de Classificadores - Ex. Médicos

- Comitê de Médicos Especialistas

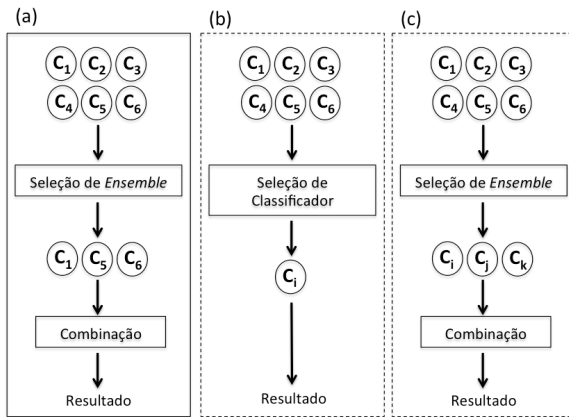


Seleção Estática x Dinâmica

- Seleção Estática
 - As regiões de competência de cada classificador são definidas durante a fase de treinamento, antes da classificação¹.
 - O mesmo classificador ou ensemble para todos os padrões de teste
- Seleção Dinâmica
 - A escolha de um classificador para o padrão de teste é feita durante a classificação. O classificador com a maior competência é escolhido¹
 - A cada padrão de teste, uma nova seleção é executada
- Problemas diferentes, soluções diferentes!

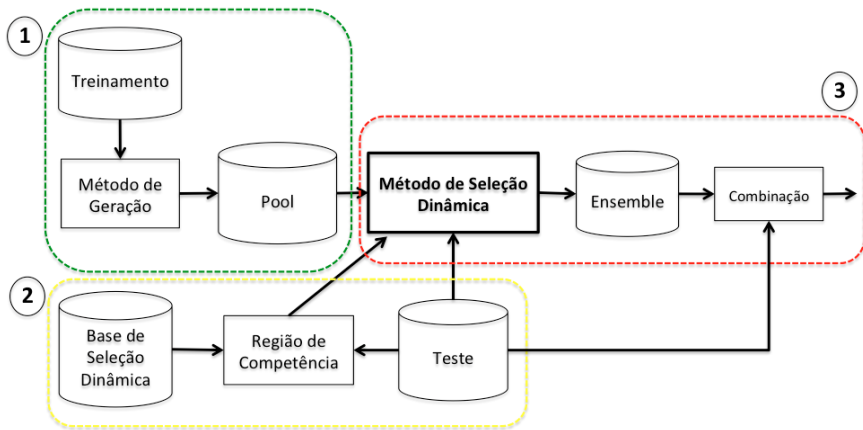
Seleção Estática x Dinâmica

- Três formas diferentes de seleção e combinação de classificadores [6]



- (a) Seleção Estática de *Ensemble*
(b) Seleção Dinâmica de Classificador
(c) Seleção Dinâmica de *Ensemble*

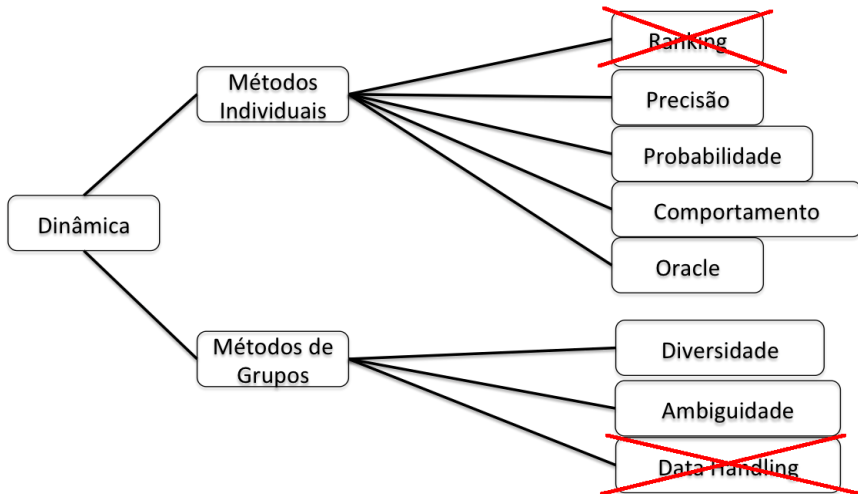
Arquitetura Básica



Métodos de Seleção Dinâmica²

- Métodos Individuais
 - Analisa o desempenho individual de cada classificador no pool
- Métodos de Grupo
 - Medidas que levam em consideração a interação entre os classificadores do pool

Métodos de Seleção Dinâmica

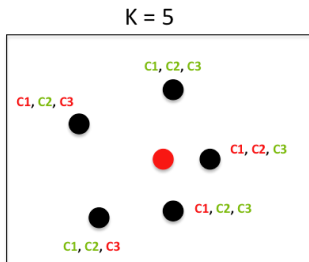


Precisão (Accuracy)³

- Estimativa da precisão do classificador global ou local
- Precisão = % dos exemplos corretamente classificados
- OLA - *Overall Local Accuracy*
- LCA - *Local Class Accuracy*

Precisão - OLA

- 1 - Submete o padrão desconhecido a todos os classificadores, se eles concordarem, retorna a resposta
- 2 - Caso contrário, calcula o % dos exemplos corretamente classificados em uma região local próxima ao padrão desconhecido



Resultado:

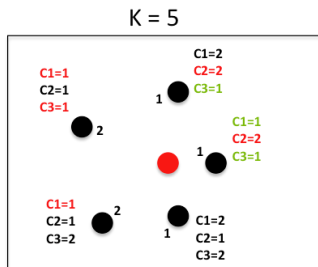
$$C1 = 2/5 = 0,4 = 3^\circ$$

$$C2 = 4/5 = 0,8 = 1^\circ$$

$$C3 = 3/5 = 0,6 = 2^\circ$$

Precisão - LCA

- 1 - Submete o padrão desconhecido a todos os classificadores, se eles concordarem, retorna a resposta
- 2 - Para o cálculo da precisão, considera apenas os exemplos que tiveram a mesma classificação do padrão desconhecido



$$C1(\bullet) = 1$$

$$C2(\bullet) = 2$$

$$C3(\bullet) = 1$$

Resultado:

$$C1 = 1/3 = 0,33 = 2^\circ$$

$$C2 = 0/2 = 0,0 = 3^\circ$$

$$C3 = 2/3 = 0,67 = 1^\circ$$

Probabilidade⁴

- Métodos que usam representação probabilística
- Dois métodos interessantes:
 - A Priori
 - A Posteriori
- Classificador selecionado baseado na sua precisão probabilística na região de competência
- A probabilidade a priori/posteriori do classificador é calculada na vizinhança do padrão desconhecido

Probabilidade

- A Priori

$$\hat{p}(\text{correct}_j) = \frac{\sum_{i=1}^K \hat{P}_j(\omega_l | \psi_i \in \omega_l) \cdot \delta_i}{\sum_{i=1}^K \delta_i} \quad (1)$$

- A Posteriori

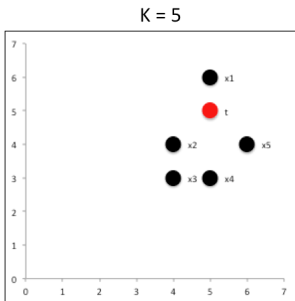
$$\hat{p}(\text{correct}_j | c_j(t) = \omega_l) = \frac{\sum_{\psi_i \in \omega_l} \hat{P}_j(\omega_l | \psi_i) \cdot \delta_i}{\sum_{i=1}^K \hat{P}_j(\omega_l | \psi_i) \cdot \delta_i} \quad (2)$$

- Peso

$$\delta_i = \frac{1}{\text{EuclidianDistance}(\psi_i, t)} \quad (3)$$

Probabilidade

● **A Priori:** $\hat{p}(\text{correct}_j) = \frac{\sum_{i=1}^K \hat{P}_j(\omega_l | \psi_i \in \omega_l) \cdot \delta_i}{\sum_{i=1}^K \delta_i}$

Classificador c_j

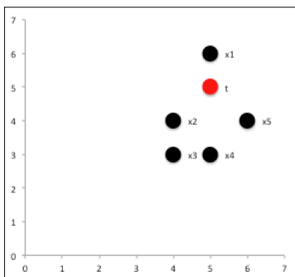
Vizinho	Classe Correta	$P_j(\omega_l / \psi_i \in \omega_l)$	δ_i	$P_j(\omega_l / \psi_i \in \omega_l) \cdot \delta_i$
x ₁	1	0.7	1	0.7
x ₂	1	0.3	0.71	0.213
x ₃	2	0.4	0.45	0.18
x ₄	2	0.8	0.5	0.4
x ₅	1	0.65	0.71	0.4615
SOMA	-	-	3.37	1.9545

$$p(\text{correct}_j) = 1.9545 / 3.37 = \mathbf{0.58}$$

Probabilidade

- **A Posteriori:** $\hat{p}(\text{correct}_j | c_j(t) = \omega_l) = \frac{\sum_{\psi_i \in \omega_l} \hat{P}_j(\omega_l | \psi_i) \cdot \delta_i}{\sum_{i=1}^K \hat{P}_j(\omega_l | \psi_i) \cdot \delta_i}$

K = 5



Classificador: $c_j(\bullet) = 2$

Vizinho	Classe Correta	$P_j(\omega/\psi_i)$	δ_i	$P_j(\omega/\psi_i) \cdot \delta_i$
x_1	1	0.3	1	0.3
x_2	1	0.7	0.71	0.497
x_3	2	0.4	0.45	0.18
x_4	2	0.8	0.5	0.4
x_5	1	0.35	0.71	0.2485
SOMA	-	2.55	-	1.6255

$$p(\text{correct}_j / c_j(t) = \omega_l) = 0.58 / 1.6255 = \mathbf{0.357}$$

Probabilidade - Algoritmo: Método A Priori/A Posteriori

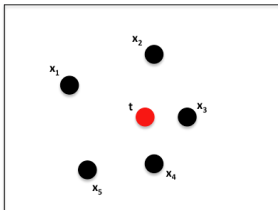
```
1: for para cada  $t$  em  $T_e$  do
2:   Encontre  $\psi$  como sendo os  $K$  vizinhos de  $t$  em  $T_r$ 
3:   for para cada classificador  $c_j$  em  $C$  do
4:     Calcular  $\hat{p}(\text{correct}_j)$  em  $\psi$ 
5:     if  $\hat{p}(\text{correct}_j) \geq 0.5$  then
6:       Guardar  $c_j$  em CS;
7:     end if
8:   end for
9:    $\hat{p}(\text{correct}_m) = \max_j(\hat{p}(\text{correct}_j))$ ;
10:   $c_m = \operatorname{argmax}_j(\hat{p}(\text{correct}_j))$ ;
11:   $\text{selecionado} = \text{TRUE}$ ;
12:  for para cada  $c_j$  em CS do
13:     $d = \hat{p}(\text{correct}_m) - \hat{p}(\text{correct}_j)$ ;
14:    if  $((j \neq m) \text{ e } (d < \text{Threshold}))$  then
15:       $\text{selecionado} = \text{FALSE}$ ;
16:    end if
17:  end for
18:  if  $(\text{selecionado} == \text{TRUE})$  then
19:     $c_t^* = c_m$ ;
20:  else
21:     $c_t^* = \text{qualquer } c_j \text{ em CS com } d < \text{Threshold}$ 
22:  end if
23:  Use o classificador  $c_t^*$  para classificar  $t$ ;
24: end for
```

Comportamento⁵

- DS-MCB (*Dynamic Selection - Multiple Classifier Behavior*)
- Estima o MCB usando uma função de similaridade para medir o grau de similaridade das saídas de todos os classificadores em relação ao padrão desconhecido
- Se um classificador é significativamente melhor que os outros ele é escolhido, caso contrário, todos os classificadores são combinados por voto majoritário

Comportamento

K = 5



Total de Classificadores = 5

$MCB_{x1} = \{2, 2, 2, 1, 1\}$

$MCB_{x2} = \{1, 1, 2, 2, 2\}$

$MCB_{x3} = \{1, 1, 1, 2, 2\}$

$MCB_{x4} = \{1, 2, 1, 2, 1\}$

$MCB_{x5} = \{2, 1, 2, 1, 2\}$

$MCB_t = \{1, 1, 1, 1, 2\}$

Fórmulas:

$$Sim(x1, x2) = \frac{1}{M} \sum_{i=1}^M Ti(x1, x2)$$

$$Ti(x1, x2) = \begin{cases} 1 & \text{se } Ci(x1) = Ci(x2) \\ 0 & \text{se } Ci(x1) \neq Ci(x2) \end{cases}$$

threshold = 0,5

$Sim(t, x_1) = 1/5 = 0,2$

$Sim(t, x_2) = 3/5 = 0,6$

$Sim(t, x_3) = 4/5 = 0,8$

$Sim(t, x_4) = 2/5 = 0,4$

$Sim(t, x_5) = 3/5 = 0,6$

CLA usando os resultados da vizinhança x_2 , x_3 e x_5

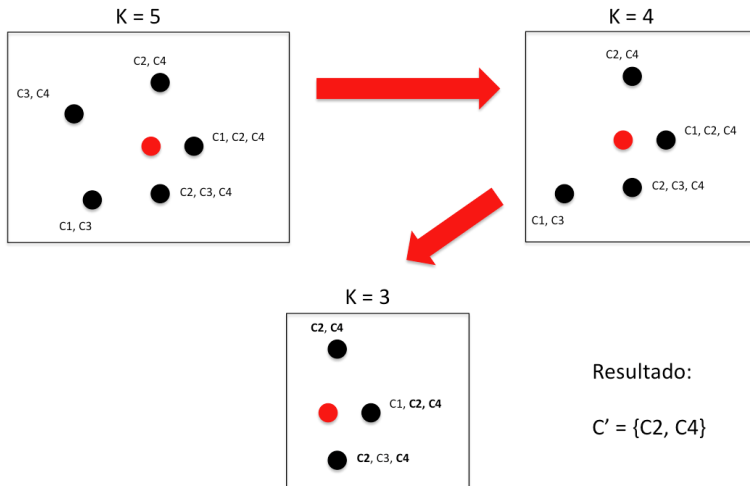
KNORA⁶ (Oracle)

- KNORA - *K-Nearest-Oracles*
- Seleciona um ensemble de classificadores mais adequado a classificar um dado padrão de teste
- Utiliza o conceito de Oracle
- Encontra uma vizinhança ao padrão de teste e quais classificadores classificam corretamente esses vizinhos
- Abordagens:
 - KNORA-E e KNORA-E-W (*Eliminate*)
 - KNORA-U e KNORA-U-W (*Union*)

KNORA-E

- Selecciona todos os classificadores que classificam corretamente todos os padrões vizinhos
- Caso nenhum o faça, diminui a vizinhança até que sobre apenas classificadores que classifiquem corretamente toda a vizinhança

KNORA-E

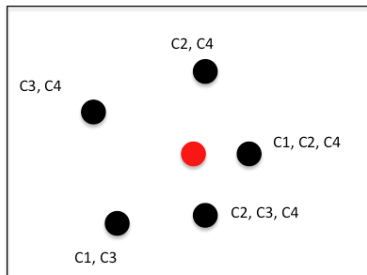


KNORA-U

- Seleciona todos os classificadores que classificam corretamente pelo menos algum padrão vizinho
- Os classificadores que tiverem mais padrões vizinhos classificados corretamente, terão maior poder de voto

KNORA-U

$K = 5$



KNORA-E-W e KNORA-U-W

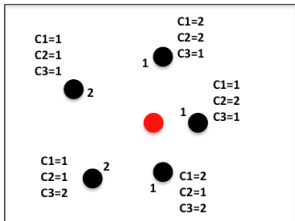
- Métodos similares aos anteriores, mas os votos são ponderados pela distância Euclidiana entre o vizinho e o padrão de teste

Diversidade⁷

- DS-KNN
 - Combina precisão com diversidade
 - Vetor 1 (decrecente) - Ordena os classificadores pela precisão
 - Vetor 2 (crescente) - Ordena os classificadores pela diversidade
 - Selecionar os classificadores mais diversos, dentre aqueles mais precisos.

DS-KNN - Diversidade

K = 5



C1(●) = 1
C2(●) = 2
C3(●) = 1

Resultado LCA:

$$C1 = 1/3 = 0,33 = 2^\circ$$

$$C2 = 0/2 = 0,0 = 3^\circ$$

$$C3 = 2/3 = 0,67 = 1^\circ$$

Resultado Diversidade – *Double Fault*

$$DF_{c1,c2} = 0,1 = 1^\circ$$

$$DF_{c1,c3} = 0,2 = 2^\circ$$

$$DF_{c2,c3} = 0,2 = 2^\circ$$

Escolha Final:

Os mais precisos: $C' = \{C3\}$

Os mais diversos, dado C' : $C'' = \{C3\}$

Ambiguidade⁸

- Diferente de Diversidade. Aqui é usado o **Consenso**
- Seleciona um ensemble dentre uma população de ensembles de alta precisão com baixa ambiguidade entre seus classificadores
- Para cada padrão de teste é selecionado de um pool de ensembles aquele que possui o maior consenso entre as saídas dos seus classificadores
- Ambiguidade de algum classificador do ensemble é determinada por:

$$a_i(\psi) = \begin{cases} 0 & \text{se } c_i(\psi) = EoC_j(\psi) \\ 1 & \text{caso contrário} \end{cases}$$

Ambiguidade

- A ambiguidade do ensemble é calculada por:

$$A = \frac{1}{|\Psi| \cdot |EoC_j|} \sum_{i \in EoC_j} \sum_{\psi \in \Psi} a_i(\psi) \quad (4)$$

- Ψ - vizinhança
- $|\Psi|$ e $|EoC_j|$ - cardinalidade dos conjuntos

Ambiguidade

Algorithm 1 Método Ambiguidade

```
1: Gera um pool de N ensembles otimizados,  $E_oC'$ , através de
   algoritmos genéticos
2: for para cada  $t$  em  $T_e$  do
3:   if todos os ensembles concordam com o rótulo de  $t$  then
4:     Classifique  $t$ ;
5:   else
6:     for para cada  $E_oC'_i$  em  $E_oC'$  do
7:       Calcule  $A_i$ ;
8:     end for
9:     Selecione  $E_oC^* = \operatorname{argmin}_i A_i$ ;
10:    Use  $E_oC^*$  para classificar  $t$ ;
11:   end if
12: end for
```

Experimentos

-Bases do UCI

-Média dos resultados para 10-Folds (8-Treinamento, 1-Validação e 1-Teste)

ÁRVORE DE DECISÃO	WINE	ECOLI	IRIS	SONAR
LOCAL HITS	96.69%	82.24%	93.33%	74.62%
OLA	89.40%	73.00%	89.33%	69.77%
LCA	89.95%	79.59%	92.00%	71.06%
KNORA-E	96.07%	77.49%	92.66%	74.22%
KNORA-U	96.69%	82.88%	92.66%	76.05%

PERCEPTRON	WINE	ECOLI	IRIS	SONAR
LOCAL HITS	95.49%	84.26%	95.33%	78.95%
OLA	96.04%	80.92%	94.00%	78.90%
LCA	95.49%	80.65%	94.66%	74.09%
KNORA-E	96.04%	79.73%	95.33%	80.90%
KNORA-U	95.49%	85.12%	96.66%	79.45%

Referências

- 1 Kuncheva, L.I., Rodriguez, J.J., *Classifier ensembles with a random linear oracle*, Knowledge and Data Engineering, IEEE Transactions on, vol. 19, no. 4, pp. 500-508, 2007.
- 2 Britto Jr, A.S., Sabourin, R., Oliveira, L.E.S., *Dynamic Selection of Classifiers - A Comprehensive Review*, Pattern Recognition, vol. 47, pp. 3665-3680, 2014.
- 3 K. Woods, K.; Kegelmeyer Jr, W.P.; Bowyer, K., *Combination of multiple classifiers using local accuracy estimates*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 4, pp. 405-410, 1997.
- 4 Giacinto, G.; Roli, F., *Methods for dynamic classifier selection*, in Image Analysis and Processing, 1999. Proceedings, 10th International Conference on, 1999, pp. 659-664.

Referências

- 5 Giacinto, G.; Roli, F., *Dynamic classifier selection based on multiple classifier behaviour*, Pattern Recognition, vol. 34, pp. 1879-1881, 2001.
- 6 Ko, A.H.R.; Sabourin, R.; Britto Jr., A.S., *From dynamic classifier selection to dynamic ensemble selection*, Pattern Recognition, vol. 41, no. 5, pp. 1718-1731, 2008.
- 7 Santana, A.; Soares, R.G.F.; Canuto, A.M.P.; de Souto, M.C.P., *A dynamic classifier selection method to build ensembles using accuracy and diversity*, in Neural Networks, 2006. SBRN '06. Ninth Brazilian Symposium on, 2006, pp. 36-41.
- 8 dos Santos, E.M.; Sabourin, R.; Maupin, P., *A dynamic overproduce-and-choose strategy for the selection of classifier ensembles*, Pattern Recognition, vol. 41, no. 10, pp. 2993-3009, Oct. 2008.