

## Combinação de Classificadores – Proba 01.

Nome: Pedro Diamel Marrero Fernández.

**QUESTÃO 1.** Na combinação dinâmica de classificadores, um pool de classificadores é gerado e os melhores classificadores são escolhidos dinamicamente para classificar o padrão de teste  $x_q$ . Sob quais condições vale a pena podar o pool de classificadores antes da etapa de seleção dinâmica dos classificadores? Justifique.

Dado um conjunto de classificadores  $H$ , o objetivo da poda é a obtenção de um subconjunto de classificadores  $E$ , sem sacrifício do rendimento global do pool inicial. Na prática os classificadores que se desejam eliminar são os classificadores que são altamente correlacionados [1]. A partir desta definição, poderíamos dizer que o uso de métodos de poda é condicionado por:

$$error(E) < error(H).$$

Ao contrário dos métodos de ensemble, os métodos de seleção dinâmica, terão que encontrar o classificador mais competente para cada padrão de teste  $xt_i$ . Esta é uma estratégia onde a eficiência do pool não é tão importante como ser capaz de obter o classificador  $h_i \in H$  que melhor classifica ao padrão  $xt_i$ . O exemplo da Fig 6.1, mostrado em [2] é uma prova disso. Neste sentido poderia pensar-se que a remoção de alguns classificadores altamente correlacionados que afetam a eficácia do pool, não beneficiará necessariamente a seleção.

Dado um conjunto de classificadores  $H = \{h_1, h_2, \dots, h_L\}$  e  $K$  regiões de competência  $R_1, R_2, \dots, R_k$ . Seja  $P(h_i|R_j)$  a probabilidade da classificação correta por  $h_i$  na região  $R_j$  em  $h_i$ . Seja  $h_{i,j} \in H$  o classificador mais competente na região  $R_j$ . Isto é:

$$P(h_{i,j}|R_j) \geq P(h_t|R_j), \forall t = 1, 2, \dots, L$$

tomemos também,

$$P(h_{i,j}|R^*) \leq P(h_t|R^*),$$

$$R^* = R - R_j$$

Isto significa que  $h_{i,j}$  é o mais competente na região  $R_j$  e o menos competente nas outras regiões. Isso tem duas consequências:  $h_{i,j}$  tem baixa precisão e pode estar altamente correlacionado com outros classificadores. Sob estas condições, este classificador seria um bom candidato para ser removido pelo algoritmo poda. Para reforçar estas hipóteses e encontrar as condições sobre as quais seria viável a utilização de poda antes da fase de seleção foram conduzidos três experimentos:

**Experimento 1).** Verificar em que medida um classificador competente para uma classe  $w_j$  é podado. Para este caso em particular foi adicionado a um pool um

classificador  $h_w$  que acerta em todos os casos para os elementos da classe  $w_j$  e erra nos outros casos. O algoritmo de poda utilizado foi EPIC. Analisou-se a posição do classificador adicionado no ordenamento realizado pelo EPIC e foi definido um erro de poda como:

$$err = \frac{h_{pos}}{L},$$

Onde  $h_{pos}$  é a posição atribuída pelo EPIC a  $h_w$ . Esse valor indica o quão longe isso está  $h_w$  de ser selecionado para formar parte do conjunto. Foi avaliada a relação da diversidade dos pools respeito aos erros obtidos. A diversidade foi calculada usando o método *Coincident Failure Diversity*.

**Experimento 2).** Tem como objetivo verificar em que medida um classificador competente para a região  $R_j$ , é podado. Para gerar um classificador competente num subespaço ou região  $R_j$ , utiliza-se o algoritmo kmeans ( $k = 10$ ) e adicionou-se ao pool um classificador  $h_r$  que acerta nos objetos que pertencem ao mesmo grupo e falha nos outros. O erro usado é o mesmo que o anterior.

**Experimento 3).** Avaliar o desempenho de vários métodos de seleção (DS-Clusters, DS-KNN, Knora-E e LCA) a partir de pools podados pelo método EPIC variando o tamanho do pool para  $L = 10, 20, \dots, 100$ .

Em todos os casos foi utilizado o banco de dados *Breast Cancer Wisconsin (Original)* da *UCI Machine Learning Repository* [3]. Como método de classificação foi utilizado *Árvore de decisão* e kfold cross-validation, para  $k = 10$ . O método *Bagging* foi usado para a geração do pool.

**Experimento 1.** Como é mostrado na Fig. 1, o erro diminui ao aumentar a quantidade de classificadores no pool. Quando os pools são mais diversos se pode observar uma queda do erro, o que significa que ao aumentar a diversidade e / ou o número de elementos do pool o classificador  $h_w$  tem mais probabilidade de ser incorporado no ensemble criado pelo EPIC. Para  $L = 50$  e  $L = 60$  encontra-se nos primeiros 5.

**Experimento 2.** Tal como no experimento anterior, as chances de pertença de um experto ao pool aumentam na medida que o pool cresce. A Fig. 2 mostra o comportamento decrescente do erro através dos diferentes tamanhos do selecionado. No entanto em dependência do por cento selecionado para a poda, há grande probabilidade de perder especialistas locais.

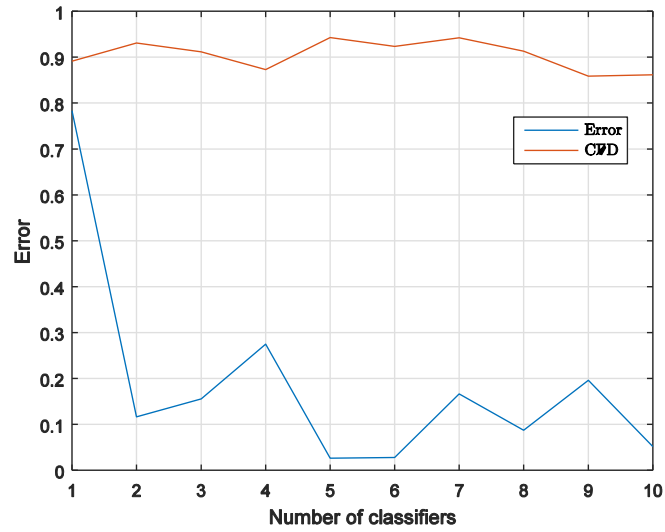


Figura 1. Resultados do erro calculado para a poda baseado na eliminação de classificadores expertos numa classe.

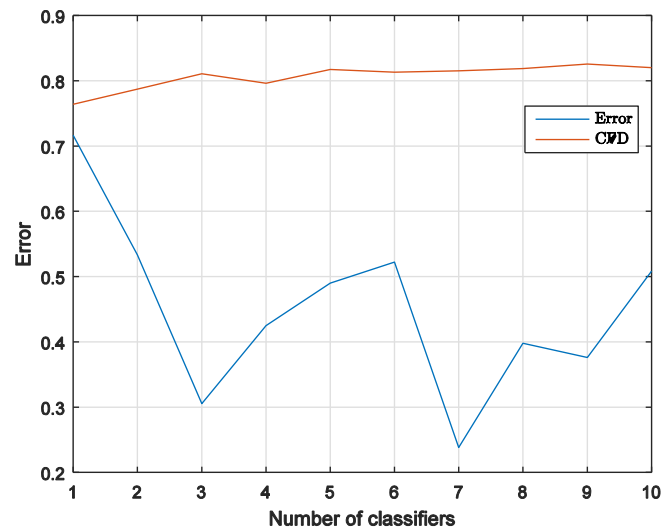


Figura 2. Resultados do erro calculado para a poda baseado na eliminação de classificadores expertos numa region.

**Experimento 3.** Como mostrado na Fig. 3, a utilização da poda não melhora o desempenho dos algoritmos de seleção utilizados. Os pools têm sido gerados com *Bagging*, no qual espera-se alta correlação entre os classificadores, característica desejada para o bom funcionamento da poda.

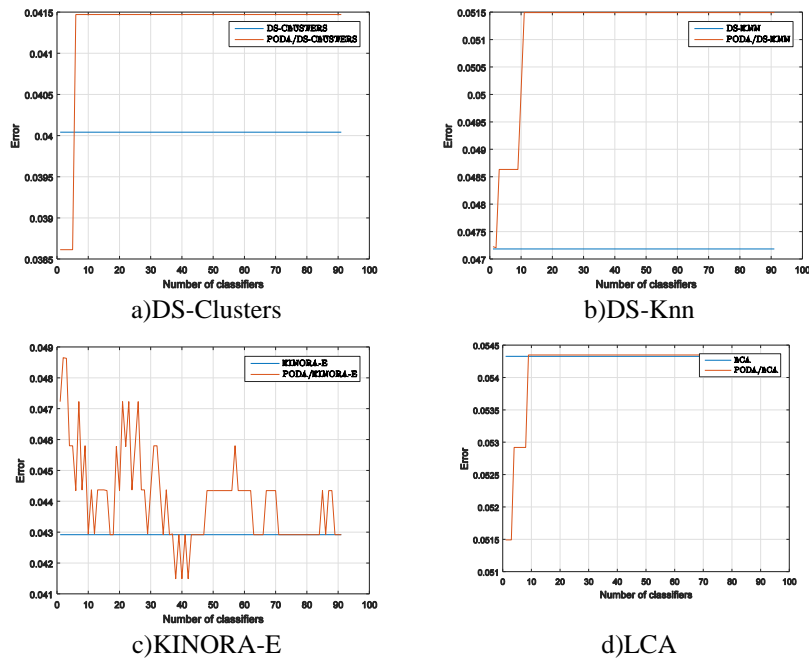


Figura 3. Resultado do erro empelando métodos de seleção para pools podados.

Como se pode observar os resultados obtidos aceitam a hipótese de que a poda pode influenciar negativamente a estratégia de seleção. Baseados nos resultados obtidos nos experimentos 1 e 2, podemos dizer que há relação entre o tamanho do pool, a correlação e os resultados da poda em relação à eliminação de especialistas. O experimento 3 mostra que os métodos de seleção utilizados neste teste não melhoram com a poda. No entanto, existem condições sob as quais eles podem ser utilizados: para os casos dos pools de elevadas dimensões e altamente correlacionados. No pior caso podem haver elementos repetidos que não contribuam à estratégia de seleção. Antes de podar deve-se garantir que a poda não afeta a estratégia de seleção através de experimentos.

**QUESTÃO 2.** Proponha uma técnicas de seleção dinâmica de ensemble, ou seja, dado um padrão de teste  $x_q$ , escolhe-se dinamicamente o melhor ensemble e os classificadores desse ensemble são combinados para predizer a classe de  $x_q$ . Discuta i) como gerar os ensembles e ii) a relação entre seleção dinâmica de ensemble e seleção dinâmica de classificadores.

A técnica de seleção dinâmica de ensemble proposta utiliza a arquitetura convencional, na qual consta de dois componentes principais, o primeiro é o componente encarregado de gerar os diferentes pools EoCs e o segundo de aplicar a seleção dos pools de EoCs dado o padrão de teste (ver Fig. 4).

Para o primeiro componente propõe-se a geração do conjunto mínimo OMC (*Oracle of Minimum Cover*) baseado na ideia do Oracle. Este conjunto, para cada elemento de

validação  $x_{val}$ , com classe  $w$  contém um subconjunto de classificadores  $OC_j$  onde todos  $h_i \in OC_j$  acertam na classificação. Estes subconjuntos serão chamados oráculos. O seguinte algoritmo mostra os passos a seguir.

---

**Algoritmo 1.** Generate Oracle of Minimun Cover

---

**Entrada:**

$C$  conjunto de classificadores.

$Val$  conjunto de validação.

**Saída:**

$OMC$  conjunto de oráculos.

**Método:**

Passo 1. Para todos os elementos do conjunto de validação  $x_{v(j)} \in Val$ .

Passo 2. Pesquisar todos os classificadores  $h_i \in C$  tal que classificam corretamente  $x_{v(j)}$  e criar o oráculo  $OC_i$ .

Passo 3. Se não existe um oráculo  $OC_j$  em  $OMC$  que classifique  $x_{v(j)}$  ( $h_i \in OC_j$  todos acertam) incluir  $OC_i$  em  $OMC$ .

---

Definiremos a probabilidade de incluir um objeto de teste  $x$ , usando um pool  $OC_i$ , à classe  $w_j$  como:

$$P(x|w_j, OC_i) = P(w_j) * \frac{P(w_j|x, OC_i)}{\sum P(w_j) * P(w_j|x, OC_i)}$$

onde

$$P(w_j|x, OC_i) = \frac{\sum (h_i(x) == w_j)}{|OC_i|}$$

Logo para classificar um objeto de teste  $x_t$ , encontram-se os oráculos do  $OMC$  dos  $k$  vizinhos de  $x_t$  no conjunto de validação e aplicamos a Regra da Soma, a partir das probabilidades dos oráculos de  $x_t$ .

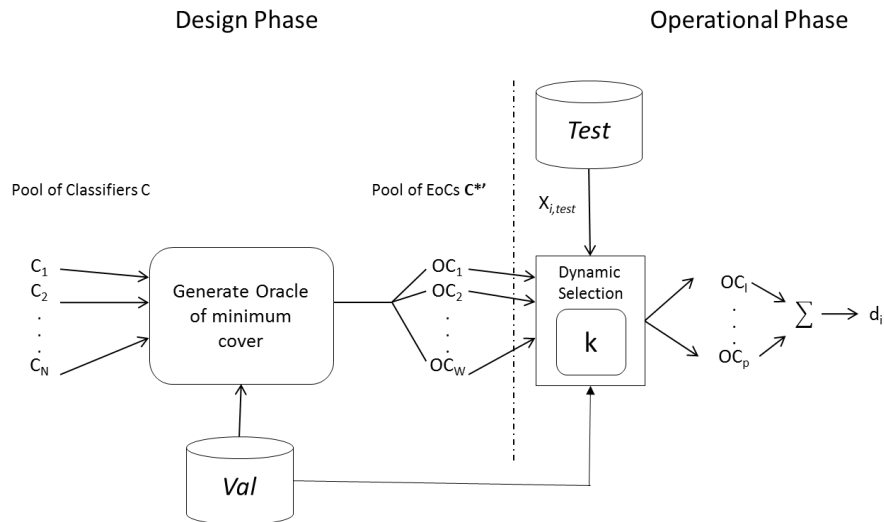


Figura 4. Esquema geral da técnica de seleção dinâmica proposta.

Um sistema de seleção dinâmica é um caso particular da seleção dinâmica de classificadores, na qual um pool de classificadores é utilizado como um classificador individual. O uso de diferentes ensembles permite combinar de maneira mais natural, diferentes tipos de classificadores e utilizar as saídas dos classificadores individuais como um novo espaço de busca (profile).

## Referencias

- [1] Zhou, Z. H. (2012). *Ensemble methods: foundations and algorithms*. CRC Press.
- [2] Kuncheva, L. I. (2004). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
- [3] Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.