

Lista de Exercício 4

Questão 1

Escolha dois métodos de seleção dinâmica, implemente e compare os seus resultados. Utilize k-fold ($k=10$) para separar os conjuntos de Treinamento (8), Teste (1) e Validação (1). Utilize para geração do pool inicial qualquer umas das técnicas aprendidas em sala de aula. Lembre-se que para cada instância do conjunto de Teste, a região de competência deve ser encontrada no conjunto de Validação e assim encontrar e combinar os classificadores.

Resposta

Os métodos de seleção dinâmica implementados foram o OLA e o LCA. A implementação foi realizada em Matlab e sua avaliação foi feita utilizando o método de validação cruzada KFold com $K = 10$, considerando 8 partições de treino, 1 partição de validação para a poda e 1 partição para teste. A base de dados escolhida foi a Ecoli do UCI Machine Learning Repository . Para os dois métodos, a região de competência é encontrada através do k-NN com $k = 5$. O número de classificadores do *pool* varia de 10 à 100 classificadores com intervalo de 10. Para cada região de competência, são selecionados 5 classificadores com melhor desempenho de acordo com a métrica utilizada pelos métodos OLA e LCA.

A figura 1 ilustra os resultados dos testes utilizando a menor taxa de erro obtida do KFold para o *pool* gerado através do Bagging e para os 5 melhores classificadores selecionados pelo OLA e LCA. A Tabela 1 mostra a média e desvio padrão das taxas de erro obtidas do KFold para o *pool* e para os classificadores selecionados pelo OLA e LCA.

Observando a figura 1, vemos que, na maioria dos casos avaliados com relação ao número de classificadores no *pool*, o menor erro dos 5 melhores classificadores selecionados pelo LCA é menor que o do *pool* gerado pelo Bagging e o dos 5 melhores classificadores selecionados pelo OLA. Na maioria dos casos, o método OLA não supera o desempenho do Bagging. Já ao se observar os valores na Tabela 1, vemos que na maior parte dos casos a média da taxa de erro do KFold e seu desvio padrão para os 5 melhores classificadores selecionados tanto pelo OLA quanto pelo LCA são menores que a média e desvio do *pool* gerado pelo Bagging, indicando, neste estudo, uma superioridade de desempenho quando se utilizam métodos de seleção dinâmica como OLA e LCA em um *pool* gerado por Bagging. A Tabela 1 também mostra que o LCA tem um desempenho superior ao do OLA, para todas as médias de erro do KFold nesse estudo.

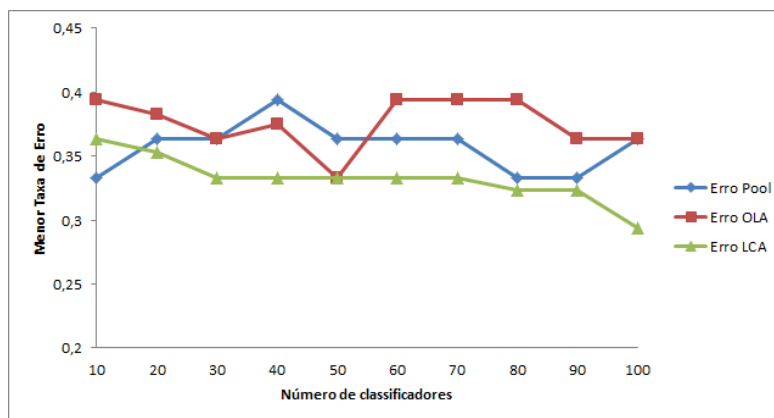


Figura 1: Resultados dos 5 melhores classificadores selecionados pelo OLA e LCA e do *pool* gerado pelo Bagging. É utilizada a menor taxa de erro do KFold com $K = 10$ e variando o número de classificadores do *pool*.

Tabela 1: Resultados da média e desvio padrão das taxas de erro obtidas pelos 5 melhores classificadores selecionados pelo OLA e LCA e pelo *pool* gerado pelo Bagging após o Kfold com $K = 10$.

Classificadores	LCA		OLA		Pool	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
10	0,416017	0,039058407	0,4515931	0,0409638	0,4576797	0,056922265
20	0,384787	0,025429395	0,4234663	0,026758203	0,4499777	0,044788375
30	0,368834	0,029451126	0,421888	0,032249923	0,4407865	0,045527462
40	0,371864	0,026835376	0,4352458	0,035214278	0,4495414	0,0516705
50	0,368997	0,025496118	0,4256443	0,051258794	0,4347259	0,04748425
60	0,371938	0,025132001	0,4457999	0,032835896	0,4485573	0,044362659
70	0,368997	0,025496118	0,4261549	0,025596297	0,439303	0,0492467
80	0,369181	0,037556803	0,4401757	0,035245871	0,4338421	0,046187651
90	0,369181	0,034903864	0,4283163	0,030867725	0,439303	0,055112792
100	0,354904	0,03245022	0,408415	0,030121549	0,4457108	0,039941771

Questão 2

No DS-KNN é utilizado um vetor de classificadores precisos, ordenado decrescentemente, e outro com uma medida de diversidade pareada para montar um vetor de diversidade crescente pelos seus valores. Seria possível, utilizando essa técnica, substituir a diversidade pareada por uma não-pareada? Como o DS-KNN poderia ser modificada para contemplar alguma técnica de diversidade não-pareada?

Resposta

Sim, é possível substituir por uma medida de diversidade não pareada. Para isso precisaríamos modificar o vetor de diversidade, utilizando grupos de classificadores ao invés de pares. Ao montar o vetor de classificadores mais precisos,

cada subconjunto de classificadores desse vetor sem repetição teria um valor de diversidade no vetor de diversidade. Então, seria selecionado o subconjunto de classificadores do vetor de precisão mais diverso. Essa escolha aumentaria o custo computacional, pois haverá mais subconjuntos sem repetição de classificadores do que pares dos mesmos.

Questão 3

As técnicas de seleção dinâmica de classificadores KNORA-E e KNORA-U produzem bons resultados comparadas a outras técnicas de DCS. Apesar do bom desempenho, existem formas de melhorar o desempenho do KNORA-E ou KNORA-U? Como? Dê ideias e, se possível, referencie utilizando artigos.

Resposta

Sim. Pode-se utilizar medidas de diversidade pareada, para reselecionar os classificadores que apresentam maior diversidade entre si. Pode-se também, para uma determinada instância de validação que tenha sido corretamente classificada por mais de um classificador, utilizar medidas de diversidade pareada para selecionar os mais diversos classificadores que acertam a classificação desta instância.