

# *Resumo*

Aqui começa o resumo escrito em língua vernácula.

# *Abstract*

Here the abstract written in foreign language begins

# ***1    Introdução***

Segundo o dicionário Cambridge, *rank* significa uma posição particular, mais alta ou mais baixa que outras.

Dado um conjunto de instâncias, espera-se obter um arranjo desse que considere alguma característica de tais instâncias.

## 2 Ranking: *Noções Preliminares*

*Ranking* pode ser enunciado, em uma forma simples, como a tarefa de ordenar os elementos de um conjunto de acordo com algum critério. Essa é uma definição bastante geral, porém estamos interessados em uma versão um pouco mais restrita do problema.

Para nossos propósitos, os elementos do conjunto a ser ordenado são denominados instâncias. Essas são compostas por um conjunto de atributos mais uma classe de domínio 0, 1.

Restringindo o problema, esperamos que o algoritmo de *ranking* receba como entrada um conjunto de instâncias no qual a classe esteja invisível durante o processo de ordenação e retorne como saída o conjunto ordenado de forma que as instâncias com classe 0 devem preceder as com classe 1.

Uma medida comum de sucesso para algoritmos de *ranking* é a área sobre a curva *ROC* (*Receiver Operating Characteristic*), comumente chamada de *AUC* (*Area Under the Curve*). A perda associada a essa medida,  $1 - AUC$ , reflete o número de instâncias, normalizado pela quantidade de 0s vezes a quantidade de 1s, que necessitam ser trocadas para um *ranking* perfeito. Para uma ordenação perfeita, em que todas as instâncias com classe 0 precedem as com classe 1, a perda na *AUC* é 0. No pior caso, em que todos os 1s precedem os 0s, a perda na *AUC* é 1.

O uso da classe como critério para ordenação sugere uma relação com os problemas de classificação e regressão em aprendizagem de máquina. Realmente, pode-se derivar uma ordenação diretamente das pontuações obtidas em uma classificação ou regressão. Se desejamos que instâncias da classe 0 ocupem as primeiras posições do *ranking*, basta ordenar de forma crescente as pontuações obtidas considerando a classe 0.

Geralmente, a medida de eficiência mais utilizada para classificação é a acurácia: uma razão entre o número de instâncias corretamente classificadas sobre o número total de instâncias no conjunto de avaliação. O erro decorrente de uma classificação afeta a acurácia de uma forma pontual.

Comparativamente, um erro de classificação pode ter maior influência na medida *AUC* que na acurácia, portanto afetar consideravelmente um *ranking*. A causa disso é a *AUC* considerar uma relação entre as instâncias, enquanto a acurácia considera apenas as instâncias pontualmente. Abaixo ilustramos através de um exemplo uma relação entre essas medidas que comprova o intuído sobre erros na classificação.

atributos	classe	previsão
$a_{0,1}, \dots, a_{0,n}$	0	0
$a_{1,1}, \dots, a_{1,n}$	0	0
$a_{2,1}, \dots, a_{2,n}$	1	0
$a_{3,1}, \dots, a_{3,n}$	0	0
$a_{4,1}, \dots, a_{4,n}$	0	0
$a_{5,1}, \dots, a_{5,n}$	0	0
$a_{6,1}, \dots, a_{6,n}$	0	0
$a_{7,1}, \dots, a_{7,n}$	0	0
$a_{8,1}, \dots, a_{8,n}$	0	0
$a_{9,1}, \dots, a_{9,n}$	1	1

Tabela 2.1: Exemplo de *ranking* e classificação

No conjunto hipotético representado ao lado, temos dez instâncias ordenadas em um *ranking* com os atributos, as classes e as previsões dadas por um classificador. Podemos perceber que o classificador errou apenas a classe da terceira instância.

Calculando a acurácia, temos nove acertos em dez possíveis, ou seja, 90%. Calculando a perda da *AUC* considerando como base a classe 0, a terceira instância precisa retroceder seis posições para uma ordenação perfeita, normalizando pelo número de 0s vezes o número de 1s, temos  $(1 - AUC) = 6 \div (2 * 8) = 0,325$ , logo a *AUC* vale 62,5%. Isso comprova que a *AUC*, comparada à acurácia, pode sofrer um impacto maior devido a erros de classificação.

Langford explica que um classificador que gere um erro de ordem  $\alpha$  na acurácia pode gerar um erro teórico máximo de  $\alpha \cdot n$  na *AUC*, onde  $n$  é a cardinalidade do conjunto de instâncias avaliado. Esse efeito se intensifica a medida que aumenta o desbalanceamento de classes do conjunto usado durante o processo de treinamento pois, quanto mais desbalanceadas as classes, mais provável que o classificador resultante seja tendencioso para a classe majoritária.

## **3     Ranking: *Implantação***

Aqui ficará o texto sobre a implantação do Ranking e as novas estratégias adotadas.

### **3.1   Votação**

Aqui ficará a explicação sobre o sistema de votação entre vários classificadores.

### **3.2   Pares por instância**

Aqui ficará a explicação sobre a quantidade de pares por instância a treinar o classificador.

## ***4    Avaliação do Ranking***

Aqui ficarão os resultados dos experimentos e as comparações entre tais.

## **5      *Conclusão***

Aqui ficará o texto da conclusão sobre o estudo dirigido.