



T1265

O IMPACTO DE ATRIBUTOS IRRELEVANTES NA PRECISÃO DOS DIFERENTES MÉTODOS DE APRENDIZADO

Henrique de Souza Oliveira (Bolsista PIBIC/CNPq e IC CNPq) e Prof. Dr. Jacques Wainer (Orientador), Instituto de Computação - IC, UNICAMP

Atributos Irrelevantes é um assunto que vem recebendo bastante atenção quando se trata de Machine Learning, pois interferem no aprendizado inserindo informações incorretas na máquina e assim reduzindo a precisão do método de classificação. Existem algumas técnicas para seleção de variáveis e remoção de atributos, porém este não é o foco deste trabalho e sim a medição de quão impactante é a adição de diversas quantidades de atributos irrelevantes ao conjunto de dados inicial. A motivação para a pesquisa nessa área é a falta de bibliografia e informações a respeito de como e quanto esses atributos afetam os diferentes métodos de classificação já desenvolvidos. Alguns estudos nesse sentido já foram realizados anteriormente, porém neles os autores analisam, não a adição de atributos aleatórios, mas sim o quanto os dados de aprendizado interferem na precisão da técnica de classificação quando escolhidos de diversas formas diferentes, isto é, qual a interferência um grupo de dados de aprendizado “desbalanceado” pode causar na classificação dos novos dados. Aqui faremos uma análise semelhante, porém ao invés de grupos de aprendizado com balanceamentos diferentes teremos grupos com diferentes números de atributos aleatórios adicionados.

Machine learning - Atributos irrelevantes - Data mining