

Ciência da Computação / Sistemas de Informação

Proposição de Técnicas de Seleção de Atributos para Problemas de Classificação Hierárquica

Mardem Arantes de Castro - 8º período de Computação, UFLA, iniciação científica remunerada

Luiz Henrique de Campos Merschmann - Orientador DAC, UFLA - Orientador(a)

Resumo

A mineração de dados, que corresponde ao processo de descoberta de padrões e conhecimentos úteis a partir de grandes volumes de dados, é extremamente importante nos dias atuais devido à grande quantidade de dados que são produzidos diariamente em todo o mundo. No entanto, ela é só uma das etapas do processo de descoberta do conhecimento, que inclui uma outra etapa que é responsável pelo pré-processamento dos dados, ou seja, por deixar os dados adequados para a etapa de mineração. Diferentes tarefas podem ser realizadas na etapa de pré-processamento dos dados, sendo a seleção de atributos uma das mais relevantes. Um dos objetivos da seleção atributos é reduzir o número de atributos originalmente existentes em uma base de dados visando a melhoria do desempenho preditivo de um modelo de classificação treinado a partir desses dados. No entanto, apesar dos benefícios da seleção de atributos para a tarefa de classificação, poucos trabalhos da literatura apresentam propostas de técnicas de seleção de atributos para o contexto de classificação hierárquica. Sendo assim, neste trabalho foi proposta a adaptação da métrica denominada Taxa de Inconsistência, utilizada para a seleção de atributos em problemas de classificação plana, para o contexto de classificação hierárquica. Essa métrica adaptada foi avaliada a partir de oito bases de dados hierárquicas de diferentes tamanhos usando um classificador hierárquico global proposto na literatura. Nos experimentos realizados, a métrica adaptada resultou em desempenhos preditivos sempre melhores ou iguais àqueles obtidos pela métrica utilizada em problemas de classificação plana, comprovando a adequação da adaptação proposta para o contexto hierárquico.

Palavras-Chave: Mineração de Dados, Classificação, Seleção de Atributos.

Instituição de Fomento: Cnpq

Link do pitch: <https://youtu.be/nA710tNjg-8>