

ANÁLISE LÓGICA DE DADOS

I Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação

Jackson de Sousa Marques Magalhaes, Tiberius de Oliveira E Bonates

Análise Lógica de Dados, do inglês Logical Analysis of Data (LAD), é um método de análise de dados que detecta padrões lógicos em conjuntos de dados e pode ser utilizado em diferentes tarefas de Aprendizado de Máquina. O objetivo deste trabalho é a produção de uma implementação do método LAD como um classificador baseado em regras para conjuntos de dados com duas classes. O método recebe um conjunto de observações, onde cada observação está associada a uma de duas possíveis classes, chamadas de “positiva” e “negativa”. Em LAD, um padrão corresponde a um subcubo não-vazio no espaço de observações que contém observações de uma única classe. Uma teoria corresponde a uma função de classificação definida em termos de padrões, responsável por atribuir uma classe a novas observações. O método LAD pode ser dividido em três etapas: pré-processamento, geração de padrões e formação de teoria. O código desenvolvido neste trabalho implementa todas essas etapas do método. O método de geração de padrões implementado é baseado na enumeração de intervalos em um espaço discreto. Uma função discriminante é utilizada na etapa de formação de teoria. Esta função consiste em uma soma ponderada dos padrões que contém a observação a ser classificada. Uma nova observação é classificada de acordo com o sinal desta soma. Nossa definição de função discriminante leva em consideração todos os padrões gerados envolvendo 2 ou 3 atributos e atribui, a cada um deles, o mesmo peso. O código também inclui um ambiente de testes, com validação cruzada do tipo hold-out. Os testes computacionais realizados mostram que o classificador implementado teve desempenho comparável ao desempenho de algoritmos de classificação frequentemente utilizados na literatura. A implementação desenvolvida neste trabalho é fornecida em repositório público a fim de estimular novos trabalhos na área. Por fim, agradecimentos ao CNPq, pela oportunidade de exercer esta bolsa de iniciação em desenvolvimento tecnológico e inovação.

Palavras-chave: Classificação. Mineração de Dados. Funções Booleanas. Aprendizado de Máquina.