Ciência da Computação / Sistemas de Informação

APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA RESOLUÇÃO DE UM PROBLEMA DE BIOINFORMÁTICA

Guilherme Henrique De Melo - a)6o módulo de computação, UFLA, iniciação científica pibic UFLA

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

Resumo

Nos últimos tempos, com os avanços tecnológicos, temos acompanhado uma grande aceleração no sequenciamento de novos genomas , aumentando a necessidade de automatização da análise dos dados gerados nesse processo. Isso tem motivado o surgimento de pesquisas voltadas para a proposição e aplicação de técnicas computacionais capazes de processar grandes quantidade de dados genômicos. Um importante passo para anotação funcional de um genoma é a identificação das sequências promotoras, as quais correspondem a segmentos do DNA que são responsáveis por sinalizar e controlar a posição onde o mecanismo de transcrição se inicia. A identificação de sequências promotoras em genomas pode ser vista como um problema de classificação onde, dado um conjunto de características de uma sequência genômica, deseja-se classificar aquela sequência como promotora ou não-promotora. Vale ressaltar que ele é considerado um problema de classificação complexo devido à natureza diversa das sequências. Portanto, neste trabalho, avaliamos a utilização de técnicas computacionais da área de mineração de dados, mais especificamente técnicas de classificação, para a resolução do problema de identificação de sequências promotoras em genomas de plantas. A tarefa de predição de promotores foi realizada com classificadores treinados a partir de bases de dados que foram pré-processadas visando a obtenção de uma quantidade reduzida de atributos para que fosse possível a obtenção de bons desempenhos preditivos com baixo custo computacional. Em experimentos preliminares de avaliação, os classificadores alcançaram uma taxa de acerto de 84% nas predições. A importância desse resultado se dá pelo fato de obtermos uma acurácia preditiva promissora utilizando apenas recursos computacionais, ou seja, com baixo custo financeiro quando comparado àquele que teríamos para realizar a classificação de sequências de DNAs a partir de experimentos in vitro.

Palavras-Chave: Mineracao de Dados, Bioinformática, Inteligência Artificial.
Instituição de Fomento: PIBIC/UFLA

Link do pitch: https://www.youtube.com/watch?v=jHqDHI4cxcI&ab_channel=GuilhermeMelo