

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe Tasaka de Melo¹, Helyane Bronoski Borges¹

¹Departamento Acadêmico de Informática
Universidade Tecnológica Federal do Paraná (UTFPR)
Rua Doutor Washington Subtil Chueire, 330, Jardim Carvalho,
CEP 84017-220 - Ponta Grossa - PR – Brasil

rafaelftmelo@outlook.com, helyane@utfpr.edu.br

Palavras-chave: Mineração de dados, Redução de dimensionalidade, Microarranjos, Autocodificadores

Resumo. Algoritmos de Aprendizagem de Máquina vem sendo cada vez mais utilizados pela sua capacidade de aprender a partir de grandes volumes de dados, como, por exemplo, dados de expressão gênica obtidos pela técnica de microarranjo. Uma característica das bases de dados de microarranjos é que, geralmente, ela é formada por grande quantidade de atributos - que correspondem aos genes - e um pequeno número de amostras. Sabe-se que dados com alta dimensionalidade podem possuir atributos redundantes e muitas vezes irrelevantes - podendo atrapalhar o processo de aprendizagem e o desempenho das predições. Métodos de redução de dimensionalidade são utilizados para reduzir a quantidade de atributos das bases de dados. Esses podem ser divididos em duas abordagens principais: a seleção e a extração de atributos. A seleção busca selecionar os atributos mais relevantes e não altera a base de dados original, já a extração possui o mesmo propósito, contudo ela altera os valores dos dados. Redes Neurais Autocodificadoras podem ser adaptadas e utilizadas para a extração de atributos e, conseqüentemente, a redução da dimensionalidade. Esta pesquisa tem como objetivo adaptar uma rede neural autocodificadora para ser utilizada na extração de atributos em bases de dados de microarranjo. Para isso, serão realizados experimentos em cinco bases de dados. Os resultados serão avaliados por meio da taxa de acerto dos classificadores: KNN, SVM, Naive Bayes e Árvore de Decisão e validados estatisticamente.