

Estudo de Seleção de Características com WEKA

Guilherme Ribeiro Olivatto¹

Thiago Ribeiro de Oliveira¹

Lúcio André de Castro Jorge²

¹Aluno de graduação em Análise e Desenvolvimento de Sistemas, Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – Campus São Carlos, São Carlos, SP, guilhermeribeiro.olivatto@gmail.com.br;

²Pesquisador da Embrapa Instrumentação, São Carlos, SP.

Nas últimas décadas os avanços tecnológicos desenvolvidos e empregados tanto em hardwares como em softwares, permitiu que as pessoas pudessem armazenar dados em quantidades cada vez maiores. Isso também se aplica a imagens digitais; por isso a importância na utilização de técnicas corretas e eficazes para a análise desses dados em especial. Visando atingir esse objetivo, técnicas de mineração de dados e de aprendizado de máquina auxiliam modelos de predição mais precisos. Muitas vezes a extração de características das imagens, nos trás algumas irrelevantes, e que podem atrapalhar o desempenho de algoritmos na construção de modelos de predição, pois não é a quantidade de características que assegura a eficiência do algoritmo. Assim, técnicas de seleção de características relevantes, se tornam essenciais como uma das etapas do pré-processamento, já que filtram a quantidade de dados que serão analisados, possibilitando uma acurácia maior e uma compreensão mais expressiva sobre os modelos gerados. Aplicando métodos de seleção de subconjunto de características conhecidos por FSS (Feature Subset Selection), podemos encontrar uma combinação específica de características relevantes aos algoritmos.

Por isso o desenvolvimento de uma ferramenta que aplique esses conceitos e alguns métodos que possibilitem fazer seleção de características a partir do histograma de uma imagem, têm grande importância para a etapa de pré-processamento das imagens que serão analisadas. Porque com isso, ocorre a diminuição do vetor de características e no tempo de processamento dos algoritmos que processarão os dados na etapa posterior. Isso se torna um filtro significativo quando existem muitas características e em grandes quantidades, a serem processadas pelos algoritmos. Após os dados serem filtrados, técnicas de mineração de dados como redes neurais e de árvore de decisão serão aplicadas pela ferramenta, tentando encontrar padrões implícitos nos dados. Tanto os algoritmos de seleção de características, como as redes neurais e os de árvores de decisão, serão utilizados da ferramenta WEKA (Waikato Environment for Knowledge Analysis), software Open Source criado e mantido pela Universidade de Waikato, Nova Zelândia. Ela conta com uma miscelânea de algoritmos e técnicas específicas para a mineração de dados.

Apoio financeiro: Embrapa.

Área: Instrumentação Agropecuária