WORCAP 2017
XVII Workshop de Computação Aplicada

Redução de Dimensionalidade de dados da Rede Social Twitter utilizando Técnicas de Aprendizado de Máquina

Marilyn Menecucci Ibanez; Marcos G. Quiles; Fernando M. Ramos

marilynmenecucciibanez@gmail.com

Na última década o crescimento do uso da internet gerou um aumento substancial na circulação de informação por meio das mídias sociais. Esse grande volume específico de dados chamou atenção da comunidade científica de diversas áreas de conhecimento que observou que um número considerável dessas informações expressam, de alguma forma, as opiniões e sentimen- tos dos usuários em relação a determinado assunto. Devido a esse grande interesse de diversas áreas da sociedade na análise desses dados, um es- tudo de melhores técnicas para a manipulação e entendimento desse tipo de dados é de grande importância para que se possa interpretar de forma rápida e precisa esse enorme volume de informação. Com base neste con- texto, este trabalho tem como objetivo realizar um estudo de classificação de sentimento usando uma representação padrão que pode ser a extraída via software WEKA - Waikato Environment for Knowledge Analysis, bag- of-words, etc, uma representação por vetor de características extraídas via técnica Autoencoder e uma representação por vetor de características ex- traídas com a técnica Deep-Autoencoder. No trabalho o WEKA será usada para avaliar a acurácia da classificação gerada com a utilização das três representações. O software WEKA possui uma coleção de algoritmos de aprendizado de máquinas que permitem a manipulação de grandes volumes de dados. A técnica de Autoenconder permite aprender uma representação para um conjunto de dados com a finalidade de redução de dimensionalidade. O DeepAutoenconder permite processar por diversas camadas ou camadas mais profundas os dados no qual se permite a redução de dimensionalidade dos dados. Os resultados obtidos com a aplicação dessas três representações possibilitará uma melhor análise de como cada representação se comporta na tarefa de redução de dimensionalidade de dados.

Mídias Sociais. Análise de Sentimento. Aprendizado de Máquina.