

A UTILIZAÇÃO DE REDES ADVERSÁRIAS GENERATIVAS PARA A DETECÇÃO DE IMAGENS FALSAS

Andressa Gomes Moreira, Luis Eduardo Araripe Gomes da Silva, Ialis Cavalcante de Paula Junior

A disseminação de notícias falsas publicadas como se fossem reais é facilitada pelo acesso em larga escala às mídias sociais. Uma das características é a manipulação em vídeos e imagens. Assim sendo, algumas técnicas de Machine Learning podem ser úteis para classificar tais informações como verdadeiras ou falsas, por exemplo, as Redes Adversárias Generativas (GANs). Nesse contexto, tal conceito foi introduzido pela primeira vez em um artigo de Ian Goodfellow e outros pesquisadores da Universidade de Montreal, em 2014. Em suma, são arquiteturas de redes neurais profundas compostas por duas redes, o gerador e o discriminador, que se opõem e melhoram gradualmente o seu desempenho. Dessa forma, durante o treinamento, a rede generativa aprende a gerar instâncias de dados falsas que se aproxima progressivamente da base de dados reais, enquanto a rede discriminadora aprende a distinguir e classificar os dados falsos e os dados verdadeiros. Logo, as Redes Adversárias Generativas estão aptas detectar a veracidade das informações em diferentes âmbitos, seja em imagens, vídeos, textos, poemas ou falas. Destarte, a rede neural será treinada para detectar se determinada imagem é falsa ou verdadeira. Portanto, para esse propósito, foi-se desenvolvidos algoritmos utilizando a linguagem de programação Python, em conjunto com a biblioteca de código aberta para aprendizado de máquina, TensorFlow, ideal para o treinamento da rede neural.

Palavras-chave: gan, imagens falsas, processamento de imagens.