

A Multimodal approach to detect Fake News in OSN's

Santosh Kumar Uppada
(COE18D005)

Department of Computer Science and Engineering
Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram

November 23, 2022

Abstract

Social Media Analytics focuses on gathering information from various Social Networks and to find insights from it. With the availability of internet, the impact of Social Networks has become high. As the problems due to Social Media is inevitable, there is a larger scope for research on mitigating the issues caused by Social Networks and its impact on the society. Major challenges of Social Networks are “Online Games and Challenges”, “Phishing”, “Cyberbullying”, “Fake News” or “Yellow Journalism”.

Fake News is defined as intentional or unintentional spread of news which is unreal or intentionally created and spread with some internal motto. Fake News is generally published in terms of text, images, audio, video, or reports. Fake news posts which contain images propagate faster through the network when compared to those posts which contain only textual content. It has been estimated that posts with images are ‘retweeted’ (shared) 11 times more than those without images. In terms of fake news in image content, it could be either tampered or used out of context. Fake images aim at impacting the psychological aspects like emotion and sentiment and influence people in propagating them. A credibility Neural Network Model (CredNN), an ensemble structure is proposed to detect fake images. CredNN is an ensemble construction of CNN architectures, with ELA technique to help identify images high digital alterations, and visual sentiment analysis to learn features that distinguish an image with negative sentiment. Weibo, a popular Chinese social network is used to evaluate the proposed model. These posts were verified by the official rumor debunking system of Weibo. The proposed model achieved an accuracy of 76.28%, and with better recall rate of 78.38% in identifying fake images.

Social posts often comes with a combination of Text, Images, Video and audio. Multimodal models are therefore required in order to analyze these posts. Often images are combined with misleading captions, termed as click-baits to enhance the reachability of the posts. There is a need to develop a framework which focuses on identifying features unique to fake news images and its corresponding captions to aid in their identification. The proposed framework works on Textual features, Visual features and Image polarity independently and combines the features from these networks using a fusion method. Fakeddit, a large-scale multimodal fake news data is used for analysis. It is observed that when BERT is used for Image caption data, Xception is used for visual features and visual sentiment data, and when these features are combined using fusion models like Concatenate and Maximum, gave better accuracy of 91.94%. The model also showed a better recall rate of 93.29%, which shows that there is less chance of fake posts being classified as real news.