AN APPROACH TO DETECT IMAGE-BASED CHARACTER ASSASSINATION ON SOCIAL MEDIA PLATFORMS USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

Sahan Udara Jayasuriya

A dissertation submitted in partial fulfilment of the requirement for
Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing
Informatics Institute of Technology, Sri Lanka
in Collaboration with
University of Westminster, UK

2020

Abstract

Character assassination and online harassment have been a major issue in social media platforms. Most of the time, the target of these attacks are celebrities and politicians. Further, there have been reports of incidents where memes of nude or intimate photos of the general public which has been released without the consent of the owner have been shared in social media platforms. Though there are options to report this kind of incident in social media platforms itself and organizations like CERT are available for support in these cases, there is no proper way to search for these images in social media sites.

A research was done in a mixed approach which includes quantitative and qualitative approaches to come up with a solution to the problem mentioned above. The primary two purposes of the research were to identify a method to detect images which are potential character assassinations and to identify the most appropriate technical methodology for the system. After the analysis, it was decided to use a combination of machine learning, image processing and text processing techniques, with the use of convolutional, recurrent and LBHS algorithms. Accordingly, a system was implemented to address the mentioned problem which contained a face recognition module, a face matching module, image classification model, two text classification models, an optical character recognition module and an image scraping module. After the implementation was completed, the built image model and text models were trained using training datasets to obtain the highest accuracy possible. Furthermore, a Sinhala language-based hate speech dataset with over 6000 records was created and published publicly as a contribution of this project.

As the ultimate result of the research project, which was conducted throughout nearly a year, a system to detect image-based character assassination on social media was developed using machine learning, image processing and text processing techniques. The system can find the images in the social media platforms which are classified as possible character attacks, which contains the given face of the end user.

Keywords:

Social media, character assassination, deep learning, image processing, text processing, face recognition, face matching, web scraping