DETECTION OF CYBERBULLYING IN IMAGES

Sruthi Madineni, Meghana Nayak, Rachana H S - Prof Aruna S
Department of Computer Science and Engineering, PES University

Problem Statement

Cyberbullying is a form of bullying which is administered using some electronic means. It has become very common with the advance of internet and usage of social media sites. This bullying can happen by various means. Posting rumours, sexual remarks, threats, disclosing a person's information without his consent being few among them.

- The purpose of this project is to build a design methodology to detect cyberbullying by using CNN model.
- This model can be integrated with applications where cyberbullying is happening to prevent users from posting certain images which might later make them prone to bullying and getting hate online.

Background

- With reference to the topic many works have been carried out on text which is widely known as sentiment analysis.
- Few works on cyberbullying in images have been done using the captions and comments underneath the image.
- Models have been designed using various feature extraction techniques like SIFT, BOVW, LLM. But this still yields low accuracy rate.

Project Requirement

System Constraints requires a high GPU memory since a lot of images have to be trained to have a better classification model. A high GPU memory affects the easy run on the local system and requires dependency on platform like Google Colab.

- Working PC, RAM > 8 GB are the hardware dependencies.
- Python, TensorFlow, Django are the software dependencies.
- Tools like Google Colab and Anaconda Navigator are required.

Design Approach

- The images in the dataset are augmented to increase the dataset size. This is done by providing different variations to the original image.
- CNN model is built using MobileNet which is a pre-trained model using Transfer Learning technique.
- Convolutional Neural Network model is built from scratch for the classification of images to bullying and non-bullying.
- CNN model is constructed using 1 convolutional layer in the beginning and later tried out with 2 and 3 convolutional layers. It has reached a saturation level in accuracy on reaching layer 3.
- Epochs is also varied to achieve a good accuracy using EarlyStopping method.
- This model is then integrated with the User Interface to show how our model can be used to detect the bullying images.
- An interface where any user having an account can upload images to their gallery is built. The model will detect if any user is trying to upload any bullying images and prevents them from doing the same.

Result and Discussions

- We have tried with different number of convolutional layer and achieved an accuracy of 89.5% for layer 1, 96.8% for layer 2 and 96.5 % for layer3.
- We have given epoch=50 in each case and used EarlyStopping function.
- Because of the early stopping the training stopped at 12,19 and 43rd Epoch respectively for layer1,2 and 3.

Summary of Project Outcome

- Using the dataset formed classification model is constructed using CNN and achieved an accuracy of 96.8% using 2 convolutional layers.
- The model is then integrated with the User Interface to show the above classification. It prevents the user from uploading any bullying image to their gallery.

Conclusion and Future Work

- Convolutional Neural Network model is built and the classification is done with ~96% accuracy.
- Future work is to detect and prevent cyberbullying in videos and audio.

References

- Hao Li-"Image analysis of cyberbullying using machine learningtechniques"-2016
- Belhassen Bayar and Matthew C. Stamm-"Design Principles of Convolutional Neural Networks forMultimediaForensics"-2017
- Shylaja S S, Abhishek Narayanan, Abhijith Venugopal, Abhishek Prasad- "Document Embedding Generation for Cyber-Aggressive Comment Detection using Supervised Machine LearningApproach"-2019.

SRUTHI MADINENI

MEGHANA NAYAK

RACHANA H S

Prof ARUNA S