

Hate Speech Detection and Prevention

Using Deep Learning

A Project Report

Submitted By

Joshua S - 2019202019

A report for the project submitted to the faculty of

INFORMATION SCIENCE AND TECHNOLOGY

In partial fulfilment for the degree of

MASTER OF COMPUTER APPLICATION



DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI-600025

APRIL 2022

Abstract

Hate speech is a subject of worry for online platforms. With powerfully expanding datasets manual mediation of posts is very inconceivable or will be tedious. Hate speech detection should be an automated task to distinguish hate speech from the provided input. In this work, a deep learning model multi-channel convolutional neural network (MCCNN) is implemented. The model consists of 3 channels of Convolutional Neural Network (CNN). Each channel is merged and connected to a fully connected layer from where the final output is obtained.

Introduction (Domain)

Deep Learning

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labelled data and neural network architectures that contain many layers.

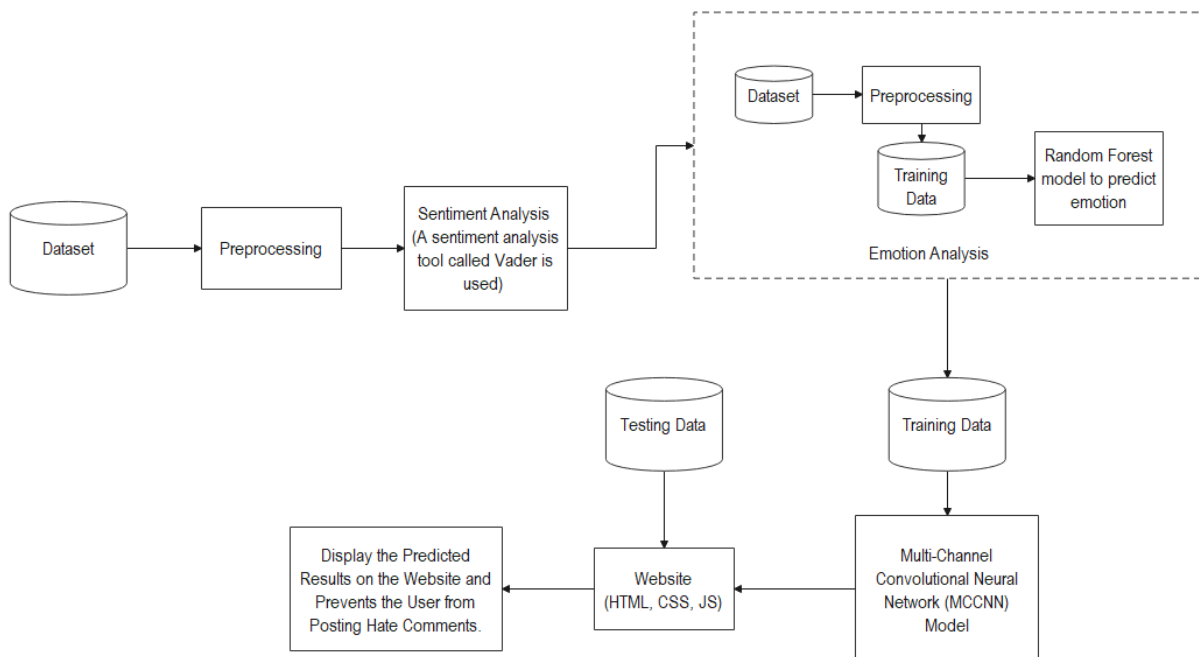
Problem Statement

Detecting the Hate Speech that has been spread online using a 3 channel Multi-Channel Convolutional Neural Network (MCCNN) which is a Deep Learning method and prevent it using suitable methods, such that a user will be restricted to post the comment online.

Objective

- To predict the Hate Speech that has been spread online
- To build an efficient model with greater accuracy & precision

Architecture Diagram



Architecture Explanation

- i. The dataset is obtained from Kaggle which contains hate tweets from twitter
- ii. The data are pre-processed by removing unused or unwanted noise from the dataset
- iii. Sentiment and Emotional analysis are performed to understand the meaning of tweets
- iv. The Data is split into training and testing for the model
- v. Multi-Channel Convolutional Neural Network (MCCNN) is used to create the model to predict Hate speech
- vi. Once the model is done, it will be connected to a website, which takes an input and classifies it as Hate speech, if there is hate speech in a statement or comment and prevents it from being posted online

List of Modules

- Data acquisition
- Pre-processing
- Sentiment & Emotion analysis
- Multi-Channel Convolutional Neural Network (MCCNN) model
- Front end

Brief Description of Modules

(i) Data acquisition:

The dataset for predicting the hate speech is obtained from Kaggle to train and test the model. The dataset consists of hate speech which is derived from tweets from twitter.

(ii) Pre-processing:

The obtained data is processed before using it to train the model. In this case, unwanted symbols, stop points, emojis, etc are removed from the dataset to reduce overfitting of the model.

(iii) Sentiment & Emotion analysis:

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral is defined as Sentiment analysis

Emotion analysis is the process of identifying and analysing the underlying emotions expressed in textual data.

(iv) Multi-Channel Convolutional Neural Network (MCCNN):

MCCNN is a CNN with more than one channel which helps the neural network to achieve more accuracy and precision.

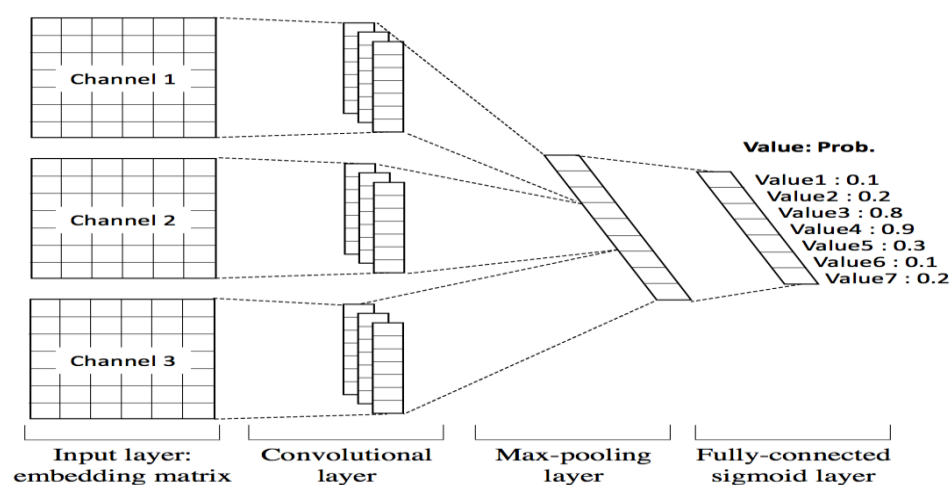


Fig. MCCNN Architecture

(v) Front end:

Finally, a website is used to visualize the hate speech detection and prevention, which is developed by HTML, CSS and JS.

References

- [A. Rodriguez, Y. -L. Chen and C. Argueta, "FADOHS: Framework for Detection and Integration of Unstructured Data of Hate Speech on Facebook Using Sentiment and Emotion Analysis," in IEEE Access, vol. 10, pp. 22400-22419, 2022, doi: 10.1109/ACCESS.2022.3151098.](#)
- [T. A. Naidu and S. Kumar, "Hate Speech Detection Using Multi-Channel Convolutional Neural Network," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking \(ICAC3N\), 2021, pp. 908-912, doi: 10.1109/ICAC3N53548.2021.9725696.](#)
- [G. H. Panchala, V. V. S Sasank, D. R. Harshitha Adidela, P. Yellamma, K. Ashesh and C. Prasad, "Hate Speech & Offensive Language Detection Using ML &NLP," 2022 4th International Conference on Smart Systems and Inventive Technology \(ICSSIT\), 2022, pp. 1262-1268, doi: 10.1109/ICSSIT53264.2022.9716417.](#)
- [N. D. T. Ruwandika and A. R. Weerasinghe, "Identification of Hate Speech in Social Media," 2018 18th International Conference on Advances in ICT for Emerging Regions \(ICTer\), 2018, pp. 273-278, doi: 10.1109/ICTER.2018.8615517.](#)

+++++