

DETECTION OF HATE SPEECH AND OFFENSIVE LANGUAGE IN TWITTER USING SENTIMENTAL ANALYSIS

ABSTRACT

The exponential growth of social media such as Twitter and community forums has revolutionised communication and content publishing, but is also increasingly exploited for the propagation of hate speech and the organisation of hate-based activities. The anonymity and mobility afforded by such media has made the breeding and spread of hate speech – eventually leading to hate crime – effortless in a virtual land scape beyond the realms of traditional law enforcement. Existing methods in the detection of hate speech primarily cast the problem as a supervised document classification task. These can be divided into two categories: one relies on manual feature engineering that are then consumed by algorithms such as SVM, Naive Bayes, and Logistic Regression (classic methods); the other represents the more recent deep learning paradigm that employs neural networks to automatically learn multi-layers of abstract features from raw data (deep learning methods). In this method we show that it is a much more challenging task, as our analysis of the language in the typical datasets shows that hate speech lacks unique, discriminative features and therefore is found in the ‘long tail’ in a dataset that is difficult to discover. We then propose Deep Neural Network structures serving as feature extractors that are particularly effective for capturing the semantics of hate speech. Our methods are evaluated on the largest collection of hate speech datasets based on Twitter, and are shown to be able to outperform state of the art by up to 6 percentage points in macro-average F1, or 9 percentage points in the more challenging case of identifying hateful content. As a proxy to quantify and compare the linguistic characteristics of hate and non-hate Tweets, we also propose to study the ‘uniqueness’ of the vocabulary for each class.

Guided by

Thirunayan Manikantan R (113216104150)

Ms.R.Amirthavalli

Aravind Kumar D (113216104016)

Assistant professor/CSE

Harish Krishna K (113216104045)

