

Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

HateSense

**Tackling ambiguity in hate speech detection – An ontology,
sentiment analysis and fuzzy logic-based approach**

A dissertation by

Mr. Keshav Kumaresan

Supervised By

Mr. Kaneeka Vidanage

Submitted in partial fulfilment of the requirements for the BEng (Hons)

Software Engineering Degree Department of Computing

BEng/BEng. (Hons) in Software Engineering

April 2019

© The copyright for this project and all its associated products resides with Informatics

Institute of Technology.

Abstract

Hate speech propagated online has been a long-trailing issue which induces several negative effects on the society. The current efforts for the automated detection of hate speech online have utilized machine learning techniques in order to try and solve the issue as a classification problem. However, the significant drawback that has been identified in existing literature is that the inability of existing systems to tackle the ambiguity when it comes to hate speech detection, more specifically differentiating between hateful and offensive content as hate speech is innately an intricate phenomenon and highly subjective from person to person. This research aims to tackle this issue of ambiguity in hopes of improving hate speech detection in general.

The proposed system will utilize human reasoning techniques such as ontologies and fuzzy logics along with sentiment analysis in order to detect hate speech and deconstruct the ambiguity present. The ontology will store key hateful terms and the magnitude of their hatefulness and provide a sense of knowledge for the detection process. The sentiment analysis modules will detect the opinions and emotions present in an input comment while the fuzzy control system will combine all the crucial factors and provide reasoning as to whether the comment can be hateful/offensive or neither. The results of the proposed approach show that the system can perform well when it comes to differentiating between hateful and offensive content and it is able to outperform existing systems in crucial factors. Yet, the deconstruction of ambiguity becomes difficult when there are a smaller number of hateful keywords present although the fuzzy control system was able to compensate in most cases. Thereby this research stresses the need for considering the disambiguation between hateful and offensive content when it comes to hate speech detection and utilization of human reasoning techniques to further facilitate this process.

Keywords: Hate Speech, Fuzzy Control, Ontology, Sentiment Analysis, Fuzzy Logic, NLP