ECE Capstone program
Spring 2018
Project Abstract & Info

Please provide the following information to be shared with on capstone information exchange platform:

1. Project Number: S18-58

2. Project Title: Toxic Comment Filtering Using Natural Language Processing

3. Team Members: Kevin Lee, Elchanan Fialkoff, Richard Santana, Peter Luo
4. Adviser(s) Name(s): Yongfeng Zhang, Kostas Bekris

5. Keywords: Natural Language Processing, Machine Learning, Neural Networks, Artificial Intelligence, Social Media

6. Project Abstract:

One does not need to look very hard to find examples of poor behavior in online discussion. With the relative anonymity provided by the Internet, more often than not, people believe that they can say whatever they want online. This sort of behavior is detrimental to the online community as a whole as it is juvenile and detracts from message boards where people could be having a real discussion.

The goal of this project is to use machine learning and natural language processing techniques to try to develop a neural network that will flag toxic comments. We define a toxic comment to be one that does not contribute to a discussion, is obscene or inflammatory, or akin to spam. We use the strategy of paragraph vectors, an extension to word embeddings where comments can be represented by a feature vector. This feature vector, once labeled, can then be used to train a neural network classifier to recognize toxicity in a comment. Due to the nature of online discussion, in which comments are often left uncorrected as far as spelling or grammar, the dataset must also be pre-processed as much as possible to minimize the chances of a comment slipping through the filter.

Through this project, we hope to gain new insights about this method of natural language processing, how it compares to other methods, and whether or not it is suitable for this purpose. It is our hope that our small contribution to the machine learning community can help make the Internet a more pleasant domain.