Allan Buchness, ahb92

Prof. Bart Selman

CS4701

7 Oct. 2020

Multiple Technique Exploration of Neural Net Classification of Offensive and Hate Speech

**INTRO**: I want to explore the use of a pre-trained neural net for semi-supervised learning on the application of detecting hate speech and offensive language. The specific application will be for a dataset of Twitter Tweets which have been classified as hate speech, offensive language, or neither split into training and test data.

**IMPLEMENTATION**: I will use neural network to implement the deep learning. First, I will secure the data and assign it randomly into two sets: training and test. Then, I will pre-process the data to a useable format for the net. This will likely involve tokenizing, case folding, removal of punctuation and non-alpha characters, possibly stemming, and possibly some treatment of stopwords. It’s also possible that I test multiple pre-processing setups to evaluate their effect on the neural network performance. Pre-processing will be part of my end-to-end model during input for training and evaluation. Then, I will convert the tokenized text data into a format digestible by the model. This involves vectorizing my text input. Finally, I will create a sequential neural net model, train the model with a percentage of the data, and then evaluate its training using the test data. Evaluation is straightforward: did it classify correctly? I will also evaluate the effectiveness in terms of runtime, and potentially space, in the case that my neural net consumes a lot of space.

**EVALUATION**: I will likely attempt to evaluate this method given different neural network setups. For example, a naïve neural network versus a pre-trained neural network so that I can compare supervised and semi-supervised learning procedures. Similarly, I will manipulate the neural network topology to evaluate the performance of different initial set-ups given the same training data. Based on my research, it seems that optimizing a neural net often depends on trial and error, so this may be my project’s focus. Perhaps I will use data outside the initial set in order to classify data which is entirely new, and I will evaluate the results as well.

**NOTES:** To accomplish these implementation goals, I plan to use Python 3.8 with the libraries: numpy, tensorflow, keras, csv, and libraries to support a potential user interface.

**TIMELINE:**

1. Oct. 20: Data Retrieval/Pre-Processing
2. Nov. 10: Vectorization and Neural Net
3. Nov. 30: Evaluation Considerations