

Emotional Analysis of Hate Speech

John Barry¹, Amber Huddell², Calien Somlak³, Dr. Randall E. Cone⁴

¹The College of New Jersey, ²Eastern University, ³Seattle University, ⁴Salisbury University

Abstract

Hate groups often propagate their messages and attract younger target audiences through subtle hate speech online. These subtle messages are very difficult to detect using standard methods of natural language processing.

Our goal is to create a parallel tokenizer alongside an emotional tagging system to test for potential patterns in the emotional content of hate speech. If there is a standard emotional pattern in hate speech, it can be used for hate speech detection. Using a large number of articles tagged with emotional content, a neural network should be able to learn these patterns if they exist.

Method

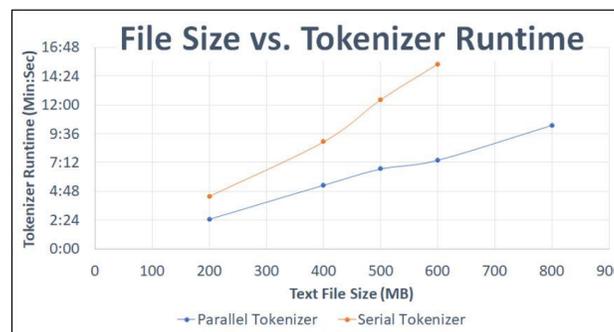
We decided to start by making a parallel tokenizer, which would help us break up text word-by-word. Then, we wanted to design a game to tag words with emotions, allowing us and future researchers to collect a larger amount of data on what emotions are related to different words. In addition, we looked at neural networks to see if we could find a pattern by hand-tagging the articles we were given.

Parallel Tokenizer

Tokenizers are a common tool in language processing that take in text and split it up into individual words for easy data manipulation. Using Python and MPI for Python, we built a tokenizer in parallel. Our tokenizer splits up the text on multiple computers at a time, then adds the words to a dictionary, mapping them to integers, reducing the space used.

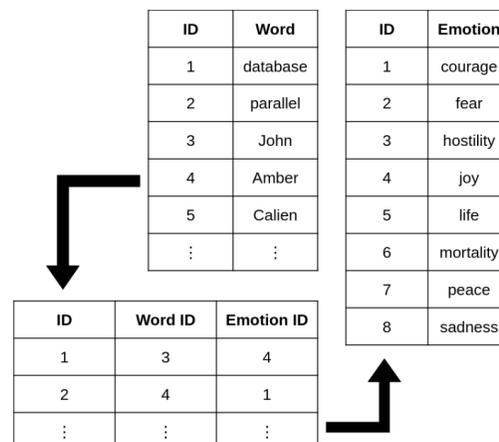
Parallel Tokenizer (cont.)

The parallel tokenizer runs in about half the amount of time as a serial tokenizer. It processes about 100MB in a minute. For comparison, the 90,000 articles we used from American Renaissance, a known hate speech website, are 161MB in total.



Emotional Tagging

After running through the tokenizer, we tagged words with emotions. The taggings are stored in an SQLite database, where words map to an ID number, which maps to an emotion ID number, which maps to an emotion. For example, "John" maps to 3, which maps to 4, which maps to "Joy." In the past, and for this project, tagging was done by hand or by using long surveys.



Gamification of Tagging

In order to speed up the tagging process and get a larger variety of opinions on what emotions relate to words, we decided to make a game. The game has a kingdom for each emotion that players get access to by paying with words they find on the ground, thereby tagging them with the kingdom's emotion. Within the kingdom, players would play minigames that would grant rewards for the player's adventure and even more data for us.



Results

After the tagging process, the word to emotion database table contained 2,081 words that were in the dictionary and tagged with an emotion. These tagged words were then used in the tagging files system which created an array of lists holding a number for every word: 0 for no emotion, or the emotional index if the word was tagged. Even after hours of work among multiple people this process showed how many words would be necessary for pattern recognition, especially after padding the data with zeros for the neural network. It achieved an accuracy of 64.6435%. This should improve as more data is provided, more words are tagged, and future work is implemented. This section cannot state anything conclusively, but this project points to a need for more data on the emotional content of words.

Conclusions

To determine if there are patterns in the emotional content of hate speech, this project focused on getting data ready for potential pattern recognition, including writing a parallel tokenizer, developing a tagging system using a database and master dictionary, and creating a game to provide data of greater quality and quantity in the future. It cannot be claimed with any evidential certainty that there exists an emotional pattern common to hate speech.

Future Work

There is still work to be done with the neural network, such as finding ways to improve the accuracy and using data from the game instead of personally-tagged words.

The game can be improved by adding more possible emotions for each word or providing ways to tag a range of emotions to a single word.

Another path for future research would be to create a network to determine whether words or phrases in and of themselves contain hateful sentiment.

Acknowledgements

This project is funded by an NSF grant for the EXERCISE REU Program. (Explore Emerging Computing in Science and Engineering)



Caroline Mansueti⁴ provided the articles used in our neural network process.