

Mr. Speice

Independent Study Mentorship 3A

7 October 2020

Understanding Depression on Twitter

Assessment 4 - Research Assessment

Date: 7 October 2020

Subject: Twitter and Depression

Works Cited:

Singh, Diveesh, and Aileen Wang. Detecting Depression Through Tweets. Stanford University, p. 9, <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6879557.pdf>.

Accessed 9 Oct. 2020.

Assessment:

To create a Twitter-based therapeutical bot for depression, it is necessary to overcome two of the most challenging aspects in natural language processing and machine learning: data collection and depression classification. Despite these challenges, researchers at Stanford University have developed data collection methods that can collect hundreds of thousands of tweets in a given moment. There have also been advancements in machine learning algorithms and architectures that can achieve up to 98 percent accuracy on detecting depression via tweets.

The first avenue explored was the data collection procedure. In Tweets with depression, there can be subtleties that might be hard for the computer to recognize. This is incredibly important because this is a very important differentiating factor in whether a tweet is classified as a depression tweet. If the algorithm used in the program does not account for the subtleties, then many tweets that likely have depression will not be accounted for. Because of how small subtleties can be, this is a challenge that requires further research. Another aspect of data

collection that was explored was keywords that led to a conglomerate of depressive tweets including but not limited to “D*mn Depression,” “Depression Notes,” or “Depression Quotes.” This serves to be important in the sense that it establishes a starting point where, in the project, we can query through tweets with these specific keywords and gather a dataset of hundreds of thousands of tweets. By having access to this many tweets, a better machine learning model can be built because there will be more training and testing data available. The final aspect that was noted was that each piece of data will be less than 280 characters, the upper character limit of a tweet. This is important to note because this means that the writing style is different in twitter than it is in a journal or book. Therefore this justifies that the data would need to be straight from Twitter rather than from a different source. Through exploring the intricacies of the data collection process, there is now an established plan for gathering and processing the data.

Another aspect that was explored was the different architectures that could be used to fulfill the actual depression analysis part. One architecture that was explored was a Recurrent Neural Network. The primary limitation of this architecture is that places more emphasis on later words than earlier words in a tweet. This is important to know because sometimes the depression part can come earlier in the tweet and the later part is irrelevant. Therefore, a recurrent neural network is likely not to be the best option to use for a classification problem. Another architecture discussed was a text-based convolutional neural network which learns word patterns and context from scratch. This would be especially beneficial to use because it would allow the product to overcome the lack of context barrier with tweets. Through evaluating the advantages and disadvantages of these architectures, there is a better understanding of how these machine learning models can be used to classify depression tweets.

In synthesis, there is a better understanding of potential nuances in data collection and depression analysis. From this, it is clear that the next steps should involve research to overcome the challenges presented by data collection and the architectures. Additionally, through this research, there is an established foundation for the steps in the development process as well and a starting point to initiate it. Collectively speaking, the analysis of data collection methods and neural network architectures has enabled further insight into the issues of straightforward development and allows us to overcome these issues by making us aware of them and allowing us to brainstorm ideas earlier within the process.

[Annotations](#)