Detecting Depression from Social Media Text

TzuHwan Seet (tseet), Sarah Branse (sbranse), Andrew Liu (aliu54), Kalvin Lam (klam4)
CSCI 1470 (Deep Learning), Department of Computer Science, Brown University. 2020

Introduction

Depression is one of the main causes of disability globablly, and suicide resulted from depression is the 2nd leading cause of death for young adults. Depression often remains undiagnosed because of social stigma, so suicide preventative programs can fail to reach people who need help. With the surge of social media use over the past decade, people are more likely to talk about mental health issues and emotions with online forums. Computational NLP methods can isolate emotions from online discussions to identify mental health cues.

The paper we chose to re-implement, *Detecting Early Onset of Depression from Social Media Text using Learned Confidence Scores* [1], seeks to develop a structural prediction model that can detect early onset of depression from users' posts on Reddit. The model we've implemented utilizes many topics we've learned about in class including word embeddings and natural language processing techniques.

Data

We utilized the eRisk (Early Detection on the Internet) dataset, which is developed from an annual workshop held by CLEF (Conference and Labs of the Evaluation Forum). The 2020 dataset has 104 depressed users and 319 non-depressed users as training data, and 40 depressed and 30 non-depressed users as testing data.

Our preprocess method sets each User with the appropriate label, depressed or non-depressed. For each post, we tokenized and stemmed the words in text and title. The users' texts are cleaned, by transforming the text into lowercase and removing punctuation and stopwords. The numbers and URLs in texts were replaced with specific tokens and then stemming was done with Porter Stemmer. Since the number of submissions from non-depressed users is so much higher than depressed users, we downsampled the majority class to a ratio of 2:1.

Methodology

We focused on the base goal of training the model to classify if a user is depressed or not. We defined our topic modelling pipeline to take in a user as input and to output a topic embedding that looks like [weight of topic 1, …, weight of topic n]. Our topic modelling pipeline determines, for each user, what topics they talk about most.

First, we represented each user as a bag of words that represented all posts from the user. We then moved onto the training phase of our Latent Semantic Indexing (LSI) Model. We trained our LSI Model by passing in all users, our dictionary we created, as well as the number of topics we chose. Once we had our LSI model trained, we printed the topics for inspection. This enabled us to see what topics the model found significant as well as what words were important to each topic. With our trained LSI model in hand, we used it to generate topic embeddings for each user. We passed in users (represented as bags of words) into our trained LSI model, and our LSI model outputted a topic embedding for each user.

These topic embeddings were then fed into the linear layer of the neural network. We used a similar architecture defined in the paper: 3 hidden layers of sizes 512, 256, and 256 respectively, using a Leaky ReLu and Dropout of 0.2. The last linear layer applied a sigmoid activation function.

Results

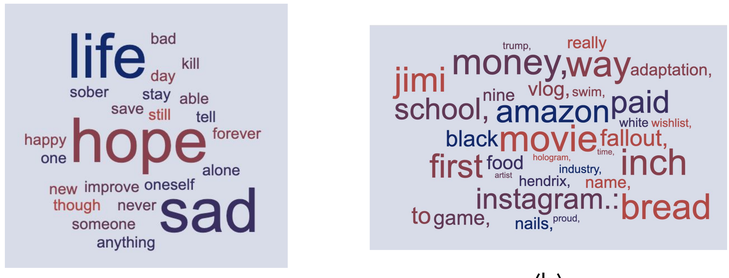
Since this a binary classification test, we evaluated many statistical measures of performance. We did not solely rely on accuracy because that metric could be misleading if the classifier tends to predict the label of the majority class (non-depressed users). As detailed by the table, our model did significantly better than the best implementation by the paper.

Table 1. Model result metrics.

Metric	Value	Scores from Paper
Accuracy	0.75	N/A
Precision	0.76	0.15-0.25
Recall	0.73	0.29-0.71
F1 Score	0.74	0.25-0.30

Discussion

Our model yielded very positive results. It scores higher on precision, recall and F1 score than the paper's best implementation. This is likely because we train the model with all of the users' posts, instead of providing the data in chunks as the original authors did to implement early detection. In order to validate these results, we also tested the model with various labels to ensure that the rate of false positives/false negatives was sensible given the inputs. In addition, we printed out topics to identify if our model successfully learned features of depressed/non-depressed users. When we explore the topics returned by the LSI model, we discover similar patterns to the results in the paper. The posts of depressed users in our dataset tend to relate to topics/themes such as *alone, depression, hugs, hopeful, happy/sad* and profanity. On the other hand, the posts of non-depressed users tend to relate to hobbies or current affairs such as *swim, school, trump, instagram.*



(a) (b)

Figure 1: Common topics for (a) depressed users and (b) non-depressed users.

Extensions

Given more time, we would definitely want to implement a model that is able to classify if the user is depressed or not with early detection confidence score. This would require us to rethink preprocessing as the user's posts need to be stored in a way that reflects the date of writing. We would also need to implement the authors' custom loss function so that the model considers the classification output only if the confidence exceeds a certain threshold.

References: [1] Bucur, A.M., Dinu, L. Detecting Early Onset of Depression from Social Media Text using Learned Confidence Scores. 2020. https://arxiv.org/pdf/2011.01695.pdf