Mental Wellness Sentiment Analysis

Jiaan Nie, 200327281

SUMMARY

The purpose of the project is to attempt to prescreen prospective clients for mental wellness issues by analyzing text. The mental wellbeing of a person is assessed through input of their text using a technique call RNN (Recurrent Neural Network). The initial plan is to create a small prototype that will work on 2 or 3 categories such as depression, anxiety and suicidal thoughts using labeled data that is scraped off from websites and organized using panda, then implement a basic RNN model using Tensorflow and test the result.

REQUIREMENTS

Hardware Requirements:

- i7 CPU
- GTX 1080 GPU

Software Requirement :

- Anaconda
- Tensorflow
- Panda
- PRAW
- WordCloud

METHODOLOGY

First scrape labeled data from reddit, use python to label the data and organize and visualize with panda.

Split the data set into:

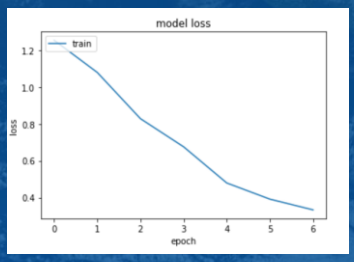
- 2/3 training
- 1/3 validation

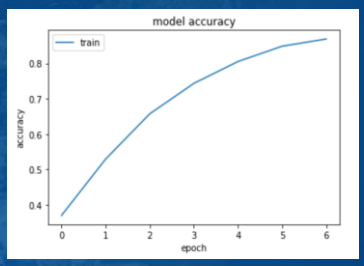
Finally feed training data into the RNN (LSTM) model and use the validation data set to access accuract.

IMPROVEMENTS

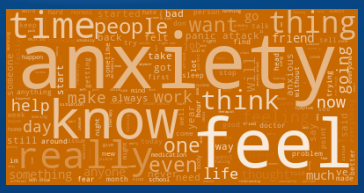
- Tuning Hyper Parameters to improve accuracy.
- Data Sanitization such as fix spelling, remove special characters , emoji etc.
- More labeled data and better quality data
- Testing on different activation functions

RESULTS





Word Clouds for the Four Different Categories



Model Summary

```
Model: "sequential_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 3149, 128)         256000
_________________________________________________________________
dropout (Dropout)            (None, 3149, 128)         0
_________________________________________________________________
lstm (LSTM)                  (None, 196)               254800
_________________________________________________________________
dense (Dense)                (None, 4)                 788
=================================================================
Total params: 511,588
Trainable params: 511,588
Non-trainable params: 0
```

Model Validation Accuracy

```
888/1 - 2s - loss: 1.2531 - accuracy: 0.5709
score: 1.26
acc: 0.57
```

CONCLUSION

Overall the accuracy of the model is not ideal, but may still be high enough to be useful when analyzed by a therapist. To improve the model will require better and more labeled data, and to sanitize the data before feeding into the model reduce noise. The model itself seems to be having trouble classifying between depression and suicidal posts due the similarity of contents. Improving these data sets may aid in classification.