ENGG 6600 (06) W24 - ST Deep Learning

Final Project Presentation

Early Detection of depression using Twitter data

Group members:

Ben Chapman-Kish

[redacted]

NOTE: This work is not exclusively my own. Only the sections I directly contributed to are included in this copy of this work.

Table of contents

01.

Introduction

What was the goal of our project?

What is the project’s topic about?

This project aims to develop a deep learning model capable of predicting the likelihood of a user developing a specific mental health disorder (depression) based on their historical tweets. By analyzing language patterns and sentiment expressed in tweets, the model can potentially identify early signs of mental health issues and facilitate timely intervention.

Background

- This concept was inspired by our desire to understand more complex NLP techniques that are made possible by advanced networks taught in this course
  - Our project specifically builds off of RNNs and LSTMs, from Week 6
- Previous works have worked with some sample users, but only mention “depression” keyword in their tweet which is not actually supporting the idea of having depression mental disorder [1]
  - Authors classified the polarity of tweets from “depressed” tweets via RNN, GRU, and CNN networks [1]

Literature Review

- Previous work includes:
  - Detecting 7 mental disorders using text and symptom features [2]
  - Classifying the severity of depression from tweets via Albert [3]
  - Classification of depressed users from their collection of tweets using CNN models [4]
  - Differentiating depressed and non-depressed tweets using a convolutional LSTM model [5]

Our Contribution

- We present a method of classifying a *user's* mental health disorder from their *history of tweets* instead of classifying *individual tweets*
- Breakdown of efforts:

  Ben: Preprocessing code and text tokenization

  [redacted]: BERT models

  [redacted] and [redacted]: LSTM/GRU models

02.

Experimental Setup and Dataset

What was our data?

Dataset

The Twitter - Self-Reported Temporally-Contextual Mental Health Diagnosis (Twitter-STMHD) dataset [6] contains eight classes of mental health disorders, each corresponding to branches in the DSM-V. These include depression, major depressive disorder (MDD), post-partum depression (PPD), post-traumatic stress disorder (PTSD), attention-deficit/hyperactivity disorder (ADHD), anxiety disorders, bipolar disorders, and obsessive-compulsive disorders (OCD). Additionally, there is a control-users class included in the dataset, totaling approximately 25,860 unique users with at least one disorder and 8000 control users.

Dataset

Dataset Structure and Collection Period:

- Users with self-reported diagnosis disclosure posts on Twitter were selected for the dataset.
- Each user's "anchor tweet", where they claim to have been diagnosed with a mental health condition, defines their data collection period.
- The collection period spans two years before the anchor tweet to two years after it, totaling a 4-year window.

Note on data ethics: the researchers did not contact any of the users to verify any diagnoses or get consent for the collection of their data in this dataset, although all of these tweets were posted publicly and this use of them is within Twitter's terms of service and user agreement.

Dataset

Anchor Tweet Identification:

- Anchor tweets were identified using a loose pattern ("diagnosed with <disorder name>") to capture tweets indicating a clinical diagnosis.
- A preliminary set of anchor tweets was collected, followed by methods to eliminate false positives.
- Two approaches were used: hand annotation and high-precision anchor tweet patterns.
- Hand annotation involved manual review of tweets to distinguish true positives, resulting in around 60% of users in the dataset.
- High-precision anchor tweet patterns were developed based on positively annotated tweets, achieving a 94% precision rate.

Anchor tweet example: *I was diagnosed with depression, anxiety, ptsd at 16. I'm now 23 and still struggle with all. But I feel like theres something else going on up in my head, has been for well over a year, its different, I feel like I know it's not any of those things, but I'm not sure what.*

Dataset

Each class of mental disorder, along with the control group, is available in its own compressed archive that takes on this structure:

Each directory represents an individual user, and from that the only file that interests us is the one containing their tweet history in the aforementioned 4-year window, `tweets.json`

We loaded the tweet JSONs and stored them in a Pandas DataFrame, initially this takes the form:

```
Python 3.12.2 (main, Feb 10 2024, 11:33:20) [Clang 15.0.0 (clang-1500.1.0.2.5)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> from preprocessing import create_tweets_df
>>> tweets_df = create_tweets_df(["depression"])
>>> print(tweets_df[["tweet_text_raw", "tweet_day", "disorder_name"]])
                                                tweet_text_raw   tweet_day disorder_name
tweet_id
1149661168908115968  Cancer research U.K. Aren't allowed to publish...  2019-07-12          None
1142024754662191104  Proud is an understatement https://t.co/FL5rxZ...  2019-06-21          None
1121607837451309056  The NHS genuinely is the jewel in our crown ht...  2019-04-26          None
1121607052814630912  I just full on had a jack and rose moment with...  2019-04-26          None
1110820069402120193                                Have mine he's a terror  2019-03-27          None
...                                                                    ...         ...           ...
1073990861409509376          *blows a kiss at the ocean(for the sharks)*  2018-12-15    depression
1073685301661331457  when the demon in ur body wants to fiesta but ...  2018-12-14    depression
1069019040000040960                              lovefool by the cardigans  2018-12-02    depression
1067525391089250304  what do u get when u cross a cowboy and a fren...  2018-11-27    depression
1066543938108964865       i wonder if u think abt me like i think abt u  2018-11-25    depression

[2116 rows x 3 columns]
```

Dataset

Sample dataset used statistics:

	Number of users	Number of tweets	disorder
depression	103	483, 618	TRUE
control	103	379, 568	FALSE

	Number of users	Number of tweets	disorder
depression	408	1,466,422	TRUE
control	408	1,752,881	FALSE

Data pre-processing

The dataset contains plenty of metadata for each tweet, such as likes and retweets, the source of the tweet (e.g. Twitter for iPhone), the timestamp, any attached media, and so on.

But primarily, the only aspect of each tweet that matters for this experiment is the tweet text itself.

Before we can start training any models on a user's tweets, we must extract from the raw tweet text only the most important aspects of it, and then vectorize that for input to our models.

Data processing scheme inspired by Garcia-Noguez et al. [4]

Text before pre-processing	Pre-processing	Text after pre-processing
HeeeeLLO Today I was diagnosed with depression ☹ which is a very common mentalillness: https://www.psychiatry.org/patients-families/depression/what-is-depression #depression	• Mentions, retweets, non-alphanumeric characters, and URLs removal • Repeated characters and words removal • Split joined words • Change emojis and emoticons to words • Stop words removal	hello today i diagnosed depression sad common mental illness depression

Data pre-processing

The process of cleaning the tweet text contains the following stages:

1. Removing URLs, retweet text, @mentions, punctuation, digits, and other special characters
2. Converting emoji and emoticons into a textual representation of what the pictogram represents
3. Removing duplicate words, repeating characters, and extra whitespace
4. Remove stop words (e.g. the, a, me, you, and, but, if, etc.)
5. Lemmatize the remaining words
    - Lemmatization is the process of reducing/converting conjugated forms or derivationally related words into one common root word (called the lemma).
    - For example: "singer", "singing", and "sang" can all be reduced to the base word "sing".

Text vectorization

After cleaning the text of each tweet, we have to perform one final transformation before the tweet is encoded in such a way that our neural networks can understand it: vectorization.

These models require well-defined numerical data as inputs, and there are a couple methods that could be used to convert text data into this. One such method is called Bag-of-Words, but for our purposes, we used the Word Embeddings method which is capable of capturing the context of a word in a tweet, semantic and syntactic similarity, its relation with other words, and so on. We used the tried-and-true Word2Vec implementation provided in the Gensim library [7] to perform this step and create our vector representation of the word embeddings in each tweet.

```
>>> vectorizer = create_word_embeddings_model(tweets_df.tweet_text_tok)
>>> tweet_embeddings = vectorizer.transform(tweets_df.tweet_text_tok)
>>> print(tweets_df.iloc[1337].tweet_text_tok)
['first', 'time', 'long', 'time', 'anxiety', 'get', 'way', 'work', 'already', 'work', 'fel
t', 'powerless', 'anything', 'maybe', 'worst', 'feel', 'weary', 'face', 'mental', 'health'
, 'month', 'mental', 'health', 'awareness', 'week', 'mental', 'health']
>>> print(tweet_embeddings[1337])
[-0.00971447  0.01548207  0.0047227  -0.00448624  0.00616144 -0.02470888
  0.0060852   0.02818106 -0.01170764 -0.00617151 -0.01219007 -0.01853303
 -0.00236404  0.00859673  0.00449295 -0.01016005  0.00380304 -0.01537228
```

Train-test split and shuffling

Once we have vectorized all the tweets and they're ready for input into a neural network, we have to split up the dataset for training and testing. We opted for a standard 80/20 train-test split.

However, we must be careful when splitting our data such that we don't test the network on tweets from a user who also had tweets in the training set. We must keep tweets from the same user together during this process. We also must stratify our data so that the proportion of depressed to control users in the original dataset is preserved in both the training and testing sets.

Only after carefully splitting and stratifying our data with these requirements in mind can we then shuffle data – except the relative order of every users' tweets must also be preserved.

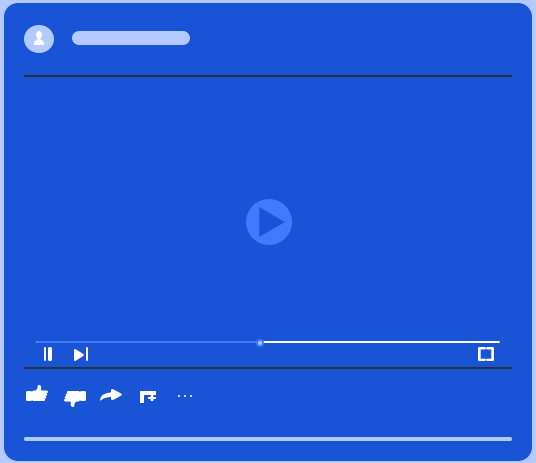
03.

Research Methodologies

Which architectures we used?

First Architecture - LSTM

1. Embedding the cleaned tweets
2. Create a sequence for each user and set max_sequence = 200
3. Padding the user's sequence less than 200
4. Adding two LSTM layer with 128 neurons
5. Adding two dense layer
6. Having binary_cross_entropy



04.

Results and Conclusion

NOTE: Results removed for public copy of this presentation.

Limitation

1. Dataset did not exclude 100% of non-English tweets
2. Diagnosing mental health disorders based solely on social media data lacks a gold standard for comparison. Without confirmed clinical diagnoses or medical records, the model's predictions may be uncertain or inaccurate. Additionally, self-reported diagnoses on social media may not always align with clinical assessments.
3. The model's performance may be affected by the evolving nature of language trends and the emergence of new slang or expressions on social media platforms. (lack of context , etc) It may also vary on a specific user's language style and cultural background.

Conclusions

- We attained a good test accuracy at predicting depression, even with our limited sample size
- Reducing the dimensionality of the inputs by cleaning and lemmatizing the tweet text both reduced the training time and enabled the models to make more accurate associations between language use and mental disorders
- There are more features available for each tweet/user from which to learn patterns behind mental disorders, but this would require a hybrid network as the data is not appropriate for use with RNNs
- LSTM performed a lot better on our learning task than GRUs with our limited data size, but a larger experiment would be needed to confirm this finding
    - Our theory to explain this is that LSTMs have more information gates and thus more trainable parameters to handle the complex dependencies between sentences as well as between entire tweets per user
    - LSTMs also take into account temporal information as well, as the timeline of when a user posted tweets makes a big difference in how we can interpret their thoughts

Reference

[1] Diveesh Singh and A. Wang, Detecting depression through Tweets - Stanford University, https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6879557.pdf (accessed Apr. 3, 2024).

[2] S. Chen, Z. Zhang, M. Wu, and K. Zhu, "Detection of Multiple Mental Disorders from Social Media with Two-Stream Psychiatric Experts," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore: Association for Computational Linguistics, 2023, pp. 9071–9084. doi: 10.18653/v1/2023.emnlp-main.562.

[3] "Depression Classification From Tweets Using Small Deep Transfer Learning Language Models | IEEE Journals & Magazine | IEEE Xplore." Accessed: Apr. 02, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9954391

[4] L. R. García-Noguez, S. Tovar-Arriaga, W. J. Paredes-García, J. M. Ramos-Arreguín, and M. A. Aceves-Fernandez, "Automatic classification of depressive users on Twitter including temporal analysis," Netw Model Anal Health Inform Bioinforma, vol. 12, no. 1, pp. 1–13, Dec. 2023, doi: 10.1007/s13721-023-00434-1.

[5] "Depressive and Non-depressive Tweets Classification using a Sequential Deep Learning Model | IEEE Conference Publication | IEEE Xplore." Accessed: Apr. 02, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10083981

[6] Suhavi, Singh, A., Singh, A. K., Shrivastava, S., Arora, U., Shah, R. R., & Kumaraguru, P. (2022). Twitter-STMHD: An Extensive User-Level Database of Multiple Mental Health Disorders. https://doi.org/10.5281/zenodo.6409736

[7] Řehůřek, R. Word2Vec Model. 2009. https://radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html

[8] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." arXiv, Feb. 29, 2020. doi: 10.48550/arXiv.1910.01108.

Thanks

Do you have any questions?