Depression Detection using twitter data

Pramil Panjawani
pramilp@iiitd.ac.in

Kriti Poddar
kritip@iiitd.ac.in

Vikas Balki
vikas17371@iiitd.ac.in

Apratim Ankit
apratim17282@iiitd.ac.in

Suverna Bisht
suverna20125@iiitd.ac.in

1. PROBLEM STATEMENT

Depression is major cause of many disease and there are majority of people(70%) who do not visit doctor for depression. Rather they rely on social sites to pour there emotions. Using computational approaches to understand mental health status (MHS) allows us to discover and identify risky behaviors, at an early stage and can be used to provide treatment and intervention, to a population that is generally hard to reach. So we try to come up with a model that can help us predict depression using social media of a person and have proper intervention at an early stage.

2. MOTIVATION

According to WHO "Depression is a common illness worldwide, with more than 264 million people affected". Same report says that "between 76% and 85% of people in low- and middle-income countries receive no treatment for their disorder". There are various effective ways in which depression can be detected clinically. But people do not report depression at an early stage due to social stigma associated with mental disorders and ignorance. People rely more on social media to pour their emotions. Thus Social media like Twitter, Facebook, reedit etc gives a great platform to users to communicate their opinions, photos and videos, which when analyzed can depict the person's moods, sentiment and feelings. Depression becomes even more difficult to distinguish because it is often misunderstood by depressed mood which can be event based like marriage, pregnancy,new job, new place etc. In this research we are trying to detect depression from the users social media activities. We will be using different features to detect depression. Through this research work we will first try to study how depression is detected clinically and then will try to devise a way to detect depression using users social media activity.

3. LITERATURE REVIEW

In the study [1],which is a survey paper of around 75 papers, we observe that most deal with binary classification problems with classes being low vs high stress, while the rest use 6 use discrete values. 25 studies do prediction per post rather than the user directly. Major work has been done by analyzing the posting behaviour and linguistic patterns of the user. Various micro blogging platform like twitter, facebook,reddit etc has been used in research..For annotating data different papers used different approaches such as expert assessments done by doctors, network affiliations, self-disclosure, selecting participants through screening questionnaires, etc.To maintain the quality of data and remove data bias, researchers used methods such as behavior threshold, removing gibberish data, post requirement, age-based restrictions, quality control surveys, removing spam posts, etc. Different papers used different features or characteristics relevant for predictions, ranging from 7 to 15,000. 68 papers used language features like syntactic features such as length, count of emojis, numbers, tagging, etc to remove the difficulties in things like word matching and semantic analysis. Word models were used to get a probabilistic distribution of characters. Topical features like latent Dirichlet allocation (LDA) and clustering have been found really effective in detection of human disorders on social media. Words used in a post related to depression, self-harm, stress, etc. were also used as domain related features. 37 papers used behavior features like posting frequency, time of activity, following, memberships, activity threads, friend groups, network, etc.38 papers used affect features valence, arousal, sentiment, and emotional intensity. It was found out that these features such as emotional were effective in solving cross model problem. 11 papers used demographic features like gender, income, education and relationship status.The different algorithms used included support vector machines, logistic regression, Random forest, decision trees, naive bias, and XGBoost for classification. But in more recent papers deep learning-based methods were more prominent.
[2] et al use deep learning to analyse social media data for early depression detection. To tackle the imbalance in real life data they propose a model X-A-BiLSTM which is the combination of XGBoost and Attention-BiLSTM. Dataset used was Reddit Self-reported Depression Diagnosis (RSDD).
[3] compare performance of various algorithms on interview data set [4] to find any patterns in spoken language, and conclude that LSTM perform the best(about 94%). [5] use MDD to establish the ground truth, from the crowd sourced data. CES-D questionnaire along with self reporting to get the labels. The users were then requested to share their twitter information for analysis. In the end they obtained 171 subjects who tested positive for depression and 305 with low or no depression. They use features such as uni-grams from depression lexicon, such as anxiety, withdrawal, suicidal, etc. Profile features used included number of followers, reciprocity graph density,etc. The final feature vector included mean, variance, momentum and entropy of number of tweets, followers, valence, dominance,words from depression lexicon, etc. They have done analysis on vari-

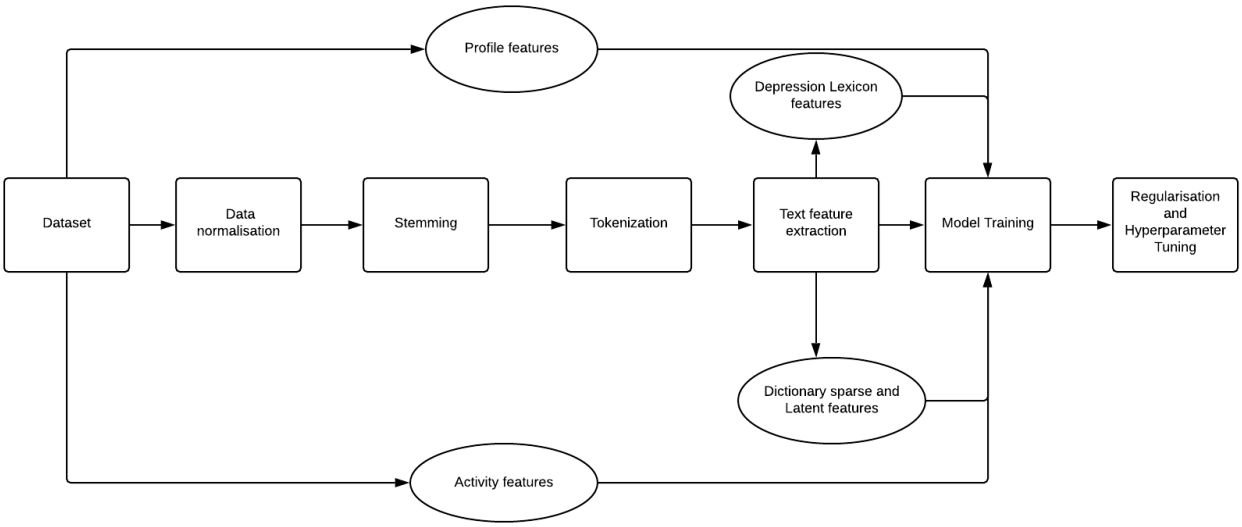


Figure 1: Plan of work

ous factors such as geography, gender and time.According to which female showed higher depressed index, depressed users where more active at late night hours and certain city showed higher depression index.[7] This paper proposed the use of different modality features to detect depression using twitter data.They have considered 6 features of the user based on the metadata of user on twitter, it's avatar etc They have analyzed the contribution of the feature modalities and detected depressed users on a twitter data set.They have analyzed different other behaviour of depressed user like dull colour usage in profile avatar, late night posting behaviour, usage of more negative words and use of first person pronoun. Using these features and behaviour analysis with Dictionary learning they were able to achieve 3% - 10 % better performance in term of F1 Score. [8]This paper showed a way to collect data automatically and then tried to check its accuracy to identify depressed groups from the control group. They have done this study for several type of mental disorders like bipolar disorder,post-traumatic-stress disorder, and seasonal effective disorder. They have shown that this data and methods are quantifiable ways to detect mental disorders. This encourage to use of social media data to detect mental disorders. [9]Automatically collected data from the twitter for the user who self reported for the these mental disorder.They have used various topic modelling to find the person interest.Basically they have compared LIWC and LDA and found out that LDA is more useful in finding the topic analysis in depression detection. [10] This paper shows that online behaviour of the user can depict the depressed state of the user. Social media helps to track the users everyday behaviour moreover since social media is ubiquitous it is easy to build real time healthcare systems for mental health. They have proved that these behaviour can be reproduced.

4. METHODOLOGY

The dataset is taken from the [6]. The authors have collected data from Twitter for different users. They consider an anchor tweet and take into account all the tweets posted by the user within one month of the anchor tweet. It consists of social media details about the users such as social network features, profile details, visual features, emotion features, topic level features and user specific features. The dataset has been divided into negative and positive data on the basis of whether the user was found to be positive or negative for depression. For a positive user, the anchor tweet contains a string in the pattern "(I'm/I was/ I am/ I've been) diagnosed depression". Users have been concluded negative for depression if the string "depress" has never been found in their tweets between 2009 and 2016. We are using a subset of this dataset. We extract and combine all the tweets posted by the users and then divide the tweets in sets of 20.

The text data is pre-processed in the following steps:

- Step 1: Convert the text to lowercase
- Step 2: Remove the user mentions from the tweets
- Step 3: Extract and remove emojis from the tweets
- Step 4: Remove links from the tweets
- Step 5: Remove punctuations
- Step 6: Perform lemmatization

Features considered:

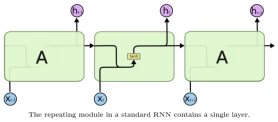
- Social Network features
- User profile
- Visual
- Emotion
- Domain-level

Further data/feature extraction steps:

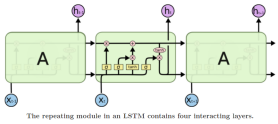
- Get users data from twitter e.g status count, follower count, etc.
- For emotion features, we take into consideration the emojis, positive word count, negative word count, first person singular count, first person plural count. The VAD features have been calculated by using a set of English words from the ANEW set.
- For extraction of visual features like brightness, contrast etc. we use OpenCV on the profile image
- We use stopwords removal, lemmatization and to match them with ANEW set to get the VAD values of the documents

4.1 LSTM

Long Short Term Memory networks (LSTMs) are a special kind of Recurrent Neural Network. They are capable of learning and remembering long-term dependencies and work well on a large variety of problems. They are able to remember information in an efficient manner. Recurrent neural networks have a chain-type formation of repeated neural network modules. In the case of a standard Recurrent neural network, the repeating module will have a simple structure, like a single tan layer.



The repeating module in a standard RNN contains a single layer.



The repeating module in an LSTM contains four interacting layers.

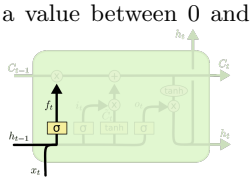
LSTM are capable of removing or adding information to the state of a cell and it's regulated carefully by structures known as gates.

LSTM have 3 types of gates:

- Forget gate
- Input gate
- Output gate

Forget Gate

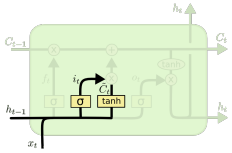
Forget Gate is used to deciding the omission of information in a particular timestamp from a cell. Sigmoid function tends to decide this information. The previous state(ht-1) and the content input(Xt) outputs a value between 0 and 1 in every cell state i.g Ct-1.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Update/Input Gate

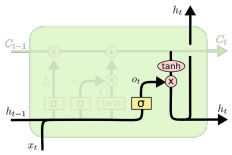
Sigmoid Functions tends to let those values 0,1. And tanh function that offers weightage to the values which are passes on basis of their importance ranging from -1 to 1.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Output Gate

Sigmoid function chooses which values to let through 0,1. also, tanh function that offers weightage to the values which are passed choosing their degree of significance going from-1 to 1 and multiplied with a yield of Sigmoid.



$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Model Implementation

We have created an LSTM model using Pytorch with torch.nn and Torch.nn.functional modules.

Following layers were defined: embedding → lstm → dropout → dense → output

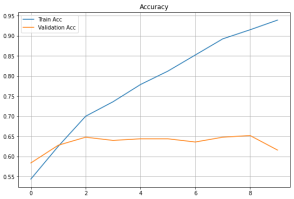
LSTM tokenizer pad the input to left up or to the right to a particular max length and if input exceeds max length then it'll truncate. This was designed so that the training for each batch can be done rather than preprocessing all inputs.

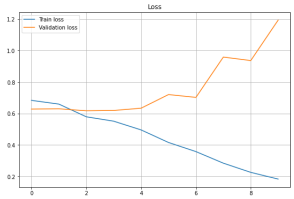
Dataset

Created a data loader for batching. There are many different ways to define them but In this, we've used a very simple solution to be used with the defined tokenizer which returns torch. tensor.

Sampling Cycle

In Input, we've taken tweet text and for the output, we have classified them into two classes: 1. Positive 2. Negative





4.2 Attention

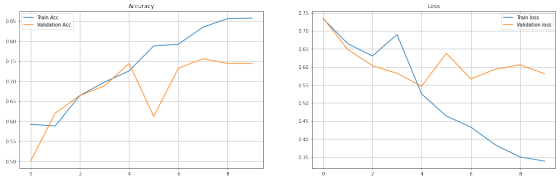
The attention mechanism emerged as an improvement over the encoder decoder-based neural machine translation system in natural language processing (NLP). The LSTM encodes the entire batch and encodes the data, the attention is used to aggregate over the entire data. This is like the summary of sentence and makes the prediction easier. To calculate this context, we use the vectors obtained from LSTM, and simply take their weighted sum. Let h_i be the hidden vector obtained from previous layers, e_{ij} be score obtained from score function **a**, w_{i-1} be weights from previous layers, and c_i be the context vector, then:

$$e_{ij} = a(w_{i-1}, h_j)$$

$$\alpha_{ij} = \exp(e_{ij}) / \sum_{k=1}^{T_x} \exp(e_{ik})$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

Now this masked output is passed through the softmax to obtain the output. We use pytorch for its implementation and with a learning rate of 0.001 and in 10 epochs we obtain accuracy of about 75%



5. FUTURE WORK

Major gaps in depression detection occur at the following places. A approach which can give a leap in this field could be using regression rather than classification. When using regression it will give a real value that will indicate the degree of depression, rather than just depressed and non depressed.

As we know that performance of ML models depends on the underline dataset. With mental health as social stigma is attached to it, it remains unreported.There is no standard dataset for depression which might be due to privacy breach of the user. So, creating a encoded dataset, which will not disclose the person's identity can be standardised and make publicly available will help researchers to focus more on the methods to detect depression.

Clinical depression is detected based on questionnaires and based on past medical history and personal life events of the user. Work can be done to extract these information from the timeline of the users. Using these features can give a better performance improvement.

While doing depression detection online we have a very wide variety of features, but correctly identifying which features depict depression is difficult. Thus efforts can be put toward identifying more relevant features on social media which can match the exact clinical process can detect depresion at its earliest. There are several disorders which are closely related to mental health like bipolar disorder,Seasonal Affecting Disorder (SAD), Psychotic Depression etc. These mental health can be detected by manipulating few features. Work can be done to deal with this type of depression also. Several features from users' profiles can be created for these issues. Mental health detection can be expanded to various other user bases based on geographical location, profession, and genders. These information can be retrieved from the metadata of the user profile. This can give a statistic to control and improvise on a person's mental health.

6. REFERENCES

[1] Chancellor, S., De Choudhury, M. Methods in predictive techniques for mental health status on social media: a critical review. npj Digit. Med. 3, 43 (2020).

[2] Cong, Q., Feng, Z., Li, F., Xiang, Y., Rao, G., Tao, C. (2018, December). XA-BiLSTM: A deep learning approach for depression detection in imbalanced data. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1624-1627). IEEE.

[3] Yuki, J. Q., Sakib, M. M. Q., Zamal, Z., Efel, S. H., Khan, M. A. (2020, July). Detecting Depression from Human Conversations. In Proceedings of the 8th International Conference on Computer and Communications Management (pp. 14-18).

[4] Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., Morency, L. P. (2014, May). The distress analysis interview corpus of human and computer interviews. In LREC (pp. 3123-3128).

[5] De Choudhury, M., Gamon, M., Counts, S., Horvitz, E. (2013, June). Predicting depression via social media. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 7, No. 1).

[6] Coppersmith G., Dredze M., and Harman C.(2014, May) Quantifying mental health signals in twitter. In The Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal To Clinical Reality, pages 51–60.

[7] Shen, Guangyao, et al. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." IJCAI. 2017.

[8] Coppersmith, Glen, Mark Dredze, and Craig Harman. "Quantifying mental health signals in Twitter." Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality. 2014.

[9] Resnik, Philip, Anderson Garron, and Rebecca Resnik. "Using topic modeling to improve prediction of neuroticism and depression in college students." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.

[10] Park, Minsu, Chiyoung Cha, and Meeyoung Cha. "Depressive moods of users portrayed in Twitter." (2012).