Depression Detection Using Tweets

Sharat Agarwal (PhD17005)

Motivation

- According to a recent report of the World Health Organization (WHO), mental health is an integral part of health and well-being.
- Mental illness is also difficult to diagnose. There is no reliable laboratory test for most forms of mental illness, typically person is diagnosed by its behaviours reported by his relatives or friends.
- Depression is very common which is so serious that by recent study around 4000 people in Canada commit suicide every year out of which 90% were diagnosed with some kind of mental disorder.
- In the world of social media people are continuously sharing their thoughts and feelings on social platforms instead of talking to their dear one personally.
- The motivation for this project is to exploit the massive data available on social platforms and use it for helping people in detecting users at risk of depression.

Problem Definition

1. Detecting Depressed/At-risk of depression users from their social media activity on twitter, by using their tweets.
2. This can be done at user-level and tweet-level
    a. Tweet-Level classification:
        i. Given a tweet, the system has to classify it as either depressed/not depressed
        ii. Due to high class imbalance(nearly 95% tweets/samples of (-ve) class), this task is very challenging.
    b. User-Level classification:
        i. Given a corpus of tweets of a certain user spanning a certain period of time. The classifier has to classify the user as depressed/not depressed.
        ii. It can leverage tweet-level classifier

Related Work

[1]. **Predicting Depression via Social Media.** Munmun De Choudhury, Michael Gamon, Scott Counts and Eric Horvitz, ICWSM AAAI Conference on Weblogs and Social Media 2013.

[2]. **Monitoring Tweets for Depression to Detect At-risk Users**, Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, and Kenton White. (CLPsych 2017), at ACL 2017.

[3] **Depression detection via harvesting social media: A multimodal dictionary learning solution.** Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, Chua TS, Zhu W. IJCAI-17

Related Work

Predicting Depression via Social Media. Munmun De Choudhury, Michael Gamon, Scott Counts and Eric Horvitz, ICWSM AAAI Conference on Weblogs and Social Media 2013.

- One of the first work in the area of Identifying Depression using Social Media.
- Used crowdsourcing to compile a set of Twitter Users who were diagnosed with Depression
    - Through their Twitter handle identified the posting behaviour of depression over the year
    - Measured the behavioral attributes relating to
        - social engagement,
        - emotion,
        - language and linguistic styles,
        - ego network, and
        - mentions of antidepressant medications
- Build a classifier using these attributes and Classify whether a person is Depressed or not.

Related Work

Monitoring Tweets for Depression to Detect At-risk Users, Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, and Kenton White. (CLPsych 2017), at ACL 2017.

- The data collected from #BellLetsTalk campaign that was multi-year program designed to break the silence around mental illness in Canada
- Used User-level Classifier and Tweet level classifier however the data is imbalanced due to data collection.
    - Used undersampling to handle class imbalance.
- Build a classifier to predict the Depressed Users using this data.

Related Work

Depression detection via harvesting social media: A multimodal dictionary learning solution. Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, Chua TS, Zhu W. IJCAI-17

- Constructed a dataset for Depressed and Non-depressed User.
- Leveraging the fact the Users use Social Media extensively to express their feelings.
- Created a Multi-modal depressive dictionary learning model to detect Depressed Users on Twitter.
- Performs better than multiple baselines by +3% to +10%

Dataset Description

1. We used the partial dataset[4] available of 5899 users classified into positive and negative classes each containing 6493 and 5384 tweets respectively.
2. Dataset contains both tweet level information and user level information.
3. Data is in raw form, with no preprocessing.

I have been diagnosed years ago with Generalized Anxiety Disorder (GAD), major depression, panic\u2026 http://t.co/idSuWjEp0c	Positive
@RebCS86 hoped you were it's just not fun. I'm good, reassured that I will never work in that horrible city \U0001f602	Negative
I have been diagnosed with major depression...I'm happy for having you as my followers I love you all...if something bad happenes	Positive
I have been sleeping so much lately. I'm probably dying... lol	Negative

Data Preprocessing

1. Stop word removal
2. Removing Punctuation
3. Convert more than 2 letter repetitions to 2 letter. Ex: funnnnny --> funny
4. Convert to lower case
5. Replaces URLs with the word URL
6. Replace @handle with the word USER_MENTION
7. Replaces #hashtag with hashtag
8. Replace emojis with either EMO_POS or EMO_NEG
9. Replace multiple spaces with a single space
10. Used porter stemmer

9 Things I Wish I’d Known About Depression Before I Was Diagnosed http://t.co/cZoc5kFc8s http://t.co/BPLPKXGmoC	*thing i wish known about depress before i was diagnos URL*

Methodology (Base Line)

1. Tf-IDF

   TF-IDF weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus.

2. Word2vec

   This model takes as input a large number of tweets and generates a vector space of typically several hundred dimensions. Each word in the corpus is being assigned a unique vector in the vector space.The powerful concept behind word2vec is that word vectors that are close to each other in the vector space represent words that are not only of the same meaning but of the same context as well.

3. Bag Of Words

   Vectorization helps in converting the text into vectors. There are two techniques

   1. Counting the number of times each word appears in a document.
   2. Calculating the frequency that each word appears in a document out of all the words in the document.

Methodology added(Emotions)

1. Positive Negative Word Count

   Using the AFINN lexicon of positive and negative words ranking from range -5 to +5, most negative to most positive word.
2. Depressed words Count

   We analyzed that depression people generally use depression related terms and drugs names more frequently in their texts.
3. Emojis Count

   Positive and negative emojis were classified and preprocessed, and their positive negative count is taken as a feature.
4. First person Singular/Plural Noun

   Use of ‘I’, ‘me’, ‘mine’, ‘my’, and plural nouns like ‘we’, ‘us’, ‘ours’, ‘our’ are more frequent in tweets by depressed people.

We combined all these features together to form a feature vector of 5 in length.

Emoticons Usage Variations and Hidden Topics :

- According to the dataset, the kind of Emoticon used are almost same
- Non-Depressed people use more emoticons than Depressed People.
- The topics that Depressed and Non-Depressed post are different

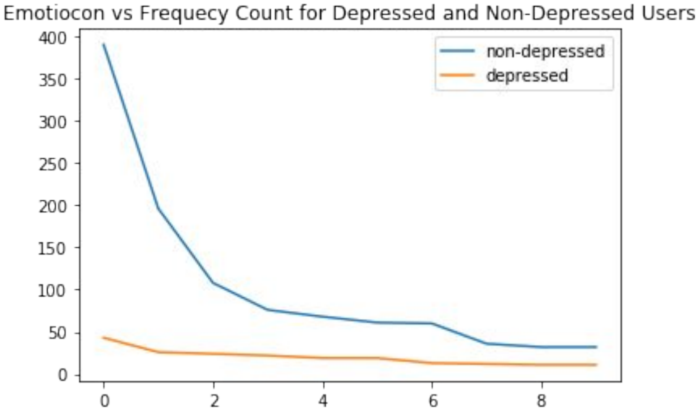
Depressed Tweets Topics:

1. Illness
2. mental
3. anxiety
4. crap

Non-Depressed Tweets Topics:

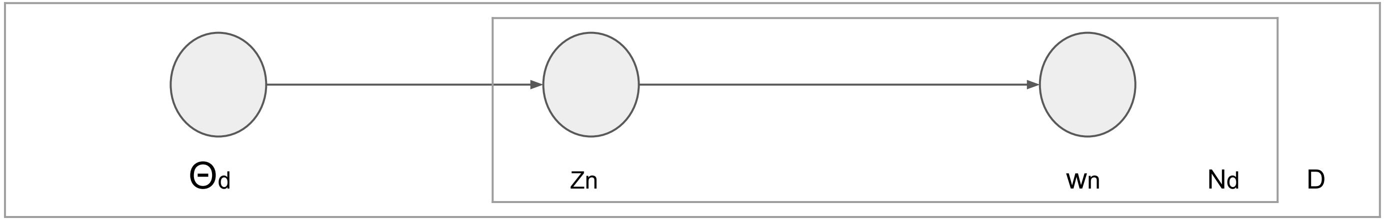
1. Sleep
2. youtube
3. think
4. love

	Non Depressed	Depressed
😂	390	43
😭	196	22
💕	32	11
❤	108	24



Methodology (LDA)

- Used LDA to extract topics from both Negative and Positive Users
  - Represents Words as Topic Vectors
  - assumes there are k underlying latent topics according to which documents are generated
    - LDA is a random distribution over unigram models p(wIΘ)
    - p(Θ;α) gives the mixture weights



Graphical model representation of LDA

- Classification with just LDA:
  - Unseen document is classified by picking argmaxcp(Clw) = argmaxcp(wlc)p(c).
- Topics extracted are taken as features for classification
  - Classification Methods includes SVM, Naive Bayes, etc.

Methodology (Gated Recurrent Unit Based Embedding)

- This is a character based approach rather than a word based approach like word2vec . In order to generate these embeddings, each tweet is processed as a stream of characters, sequentially by a GRU network.

- Our main motivation to work at character level was to achieve better generalization to Out Of Vocabulary(OOV) words. As a majority of word types feasible in any language is usually absent from the training corpus, character based approach would reduce such OOV words.

- Also, text in social media is heavily unstructured; full of slangs, abundant mis-spellings, special characters, emoticons etc. Working at character level reduces the vocabulary size as the same characters (as present in vocabulary from training corpus) is usually used to construct unseen words(OOV words).

- Also, working at character level eliminates the requirement for word segmentation, a standard Natural Language Processing(NLP) pre-processing technique. We consider white spaces as a character too.

- During pre-processing we removed the hashtags, NLTK stopwords, mentions, user names, HTTP links, retweets from all the tweets.

Operations performed inside a GRU unit.

- **Update Gate (z_t)**:

This gate controls how much past information should be propagated to future GRU units.

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$$

- **Reset Gate (r_t):**

This gate controls how much past information to forget

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$$

- **Current memory content (h'_t) :**

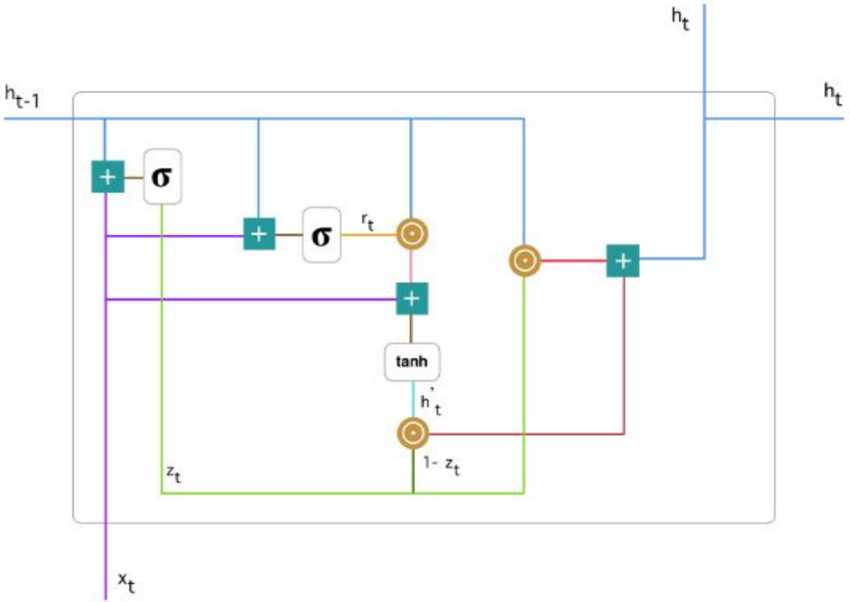
This gate represents the current memory content.

$$h'_t = \tanh(Wx_t + r_t \odot Uh_{t-1})$$

- **Final memory at current time (h_t):**

This gate represents the final output of the corresponding GRU unit.

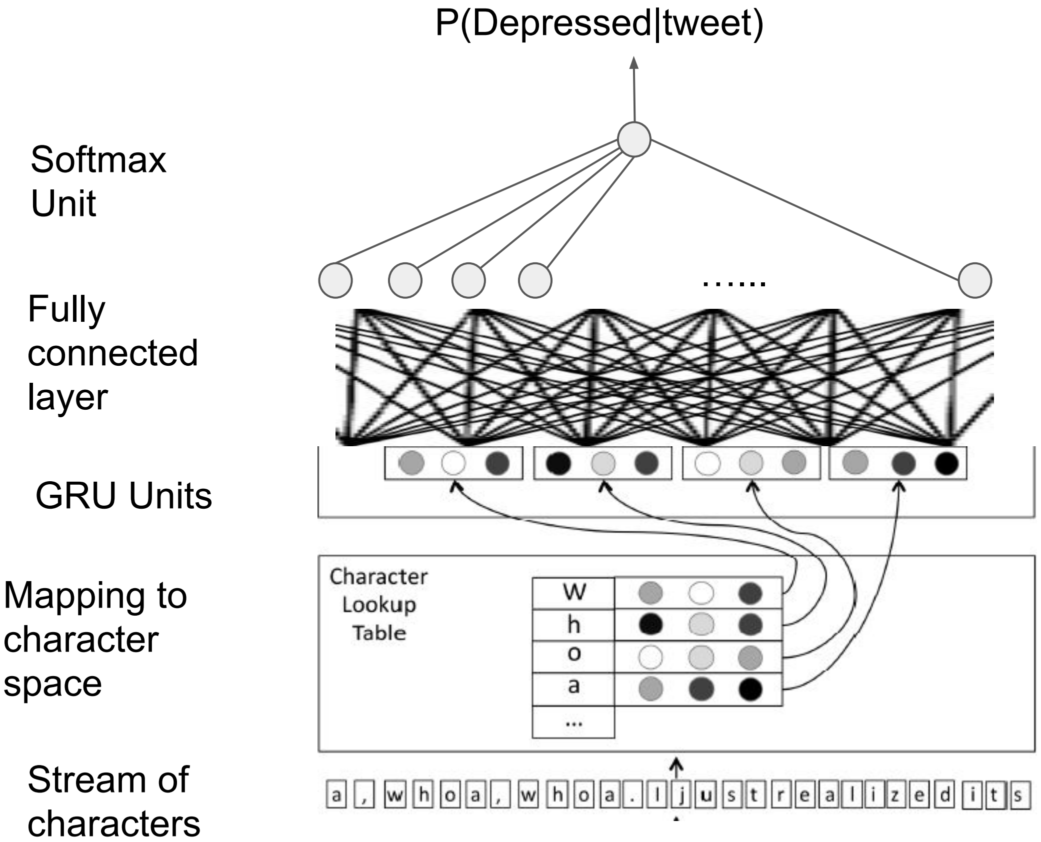
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t$$



A Gated Recurrent Unit (GRU)

PIPELINE:

- An input tweet is broken down into a stream of m characters $c_1,c_2,...,c_m$. Each of this character is one-hot encoded that is each of these is a 1x|C| vector, where C is the unicode character set obtained from the training corpus after pre-processing. It basically consists of all unique characters encountered in the training corpus.
- Each of these character vectors are then multiplied with the matrix $P_c \in R^{|C| \times d_{c}}$, where d_c represents the character space dimensions. As a result of this they get mapped to a character space and get transformed to $x_1,x_2,...,x_m$.



- The outputs of each GRU unit $h_1, h_2, \ldots, h_m$ are then fed to a fully connected layer of dimensions D which is the embedding dimension. Basically the output of this layer is the desired embedding. It is then followed by a sigmoid activation function neuron unit. The output of this unit lies between 0 and 1.

- Training of the network in order to learn all the weight parameters is done via minimizing cross entropy loss:

$$loss = \Sigma_{n=1}^{N} - t_i log(p_i) + \lambda ||\Theta||^2$$

Evaluation Metric

For evaluation we have used Precision, Recall, F1 Score and accuracy.

Accuracy in itself is not a good measure as data bit unbalanced and also we have to see which classes are misclassified.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad \text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Results

	Precision			Recall			F1 Score			Accuracy		
	LR	NB	SVM	LR	NB	SVM	LR	NB	SVM	LR	NB	SVM
Tf-IDF	0.93	0.72	0.53	0.98	0.72	0.99	**0.91**	0.842	0.692	**0.90**	0.802	0.59
BOW	0.99	0.86	0.99	0.97	0.96	0.97	0.89	0.87	**0.905**	**0.91**	0.86	0.89
Word2vec	0.86	0.79	0.81	0.91	0.81	0.86	0.85	0.81	**0.82**	**0.86**	0.78	0.81
Emotions	0.64	0.56	0.62	0.707	0.734	0.829	**0.716**	0.64	0.713	0.633	0.58	**0.67**
LDA	0.95	0.93	0.923	0.95	0.95	0.98	**0.95**	0.93	**0.95**	0.95	0.94	**0.96**
GRU	0.93	0.814	0.911	0.947	0.895	0.972	**0.942**	0.853	0.940	**0.92**	0.831	0.933

Conclusion

- We hence found out that People use Social Media very extensively to express the feelings
- Depression can be detected on the Social Media by using the way people make their post.
- Using GRU and various modalities the efficiency can be improved.

Refrences

[1]. **Predicting Depression via Social Media.** Munmun De Choudhury, Michael Gamon, Scott Counts and Eric Horvitz, ICWSM AAAI Conference on Weblogs and Social Media 2013.

[2]. **Monitoring Tweets for Depression to Detect At-risk Users**, Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, and Kenton White. (CLPsych 2017), at ACL 2017.

[3] **Depression detection via harvesting social media: A multimodal dictionary learning solution.** Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, Chua TS, Zhu W. IJCAI-17

[4] **Latent Dirichlet Allocation.** Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Advances in neural information processing systems. 2002

[5]Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, HolgerSchwenk, and Yoshua Bengio. 2014.Learning phrase representations using rnn encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Thank You