Short Summary and important points and concepts on Text-based depression detection on sparse data

Models patient's response during clinical interviews – Sparse scenario [But it can useful to us. Because in any social media platform the amount of ppl posting on suicide per day or on a time frame will be far more less compared to other posts . eg: entertainment, politics, stock market etc]

Multi task loss function -> depression severity and binary health state [depression presence or not]

Semantics and context aware -> more successful in depression detection

Depression detection[binary classification] and PHQ-8 [regression] are correlated but one can't predict the other. [Both are important]

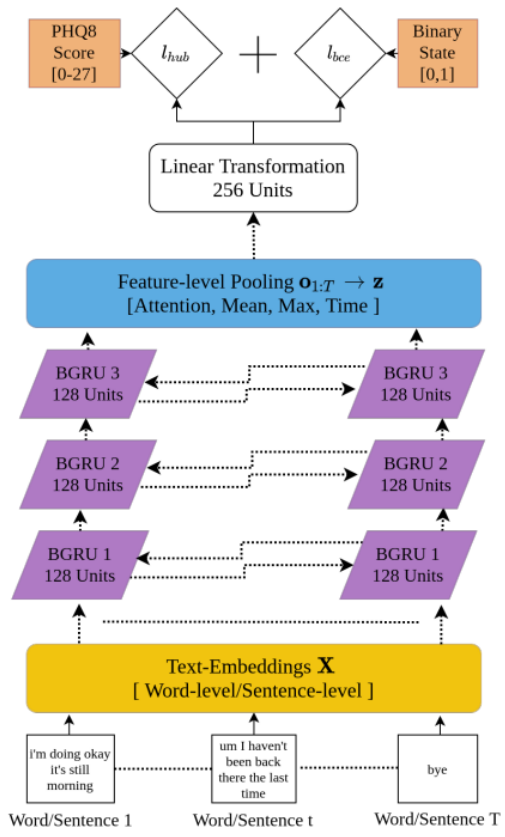
Approach in this paper:

Model -> Bidirectional Gated Recurrent Unit [Tried BLSTM but avg results were bad]

Weighted Combination of classification and regression

Classification -> Binary Cross Entropy +

Regression -> Huber loss



Sentence embedding is better to capture context

Embeddings used in this paper:

Sentence level Word2Vec/fastText embeddings

$$\mathbf{x}_j = \frac{1}{T_j} \sum_{t=1}^{T_j} \mathbf{w}_t$$

BERT [36] and ELMo [35] made use of the self-attention mechanism and LSTMs to build context-sensitive sentence representations

Dataset

- WOZDAIC (USC on distress analysis interview corpus)-> Manually labelled data from interviews (audio, video, manually transcribed text data)

Obsevations of n-grams from the above dataset

1) Very little difference in content words between healthy and depressed patients

2) Most n-grams do not contain meaningful information, e.g., "I don't know, I don't have, I do".

Data Preprocessing

Patients response without any regard to interactions with the questioner.

treats an entire paragraph spoken by a patient as a single sample

5-Fold cross-validation [80% training and 20% testing]

Neural network training

Pretrained vs From scratch

The model producing the lowest loss on the held-out crossvalidation set was chosen for evaluation on the development (dev) set

Discuss more about the selection of hyper parameters, wt and bias initializations etc

Evaluation metric

primary metrics - F1, Accuracy and MAE

secondary metrics - precision, recall and RMSE

Usually other works did thresholding the PHQ-8 regression result with a value of 10 but this work decouples classification and regression performance.

For all experiments -> report average scores after 100 runs on each fold

Findings

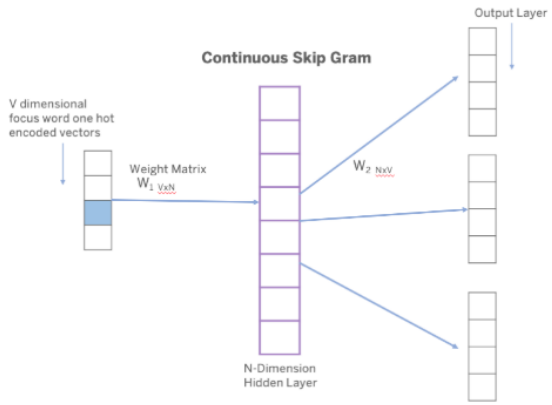
Sentence level is best compared to word level embedding

Pretrained is best compared to trained from scratch [clinical interview involves similar content to a normal conversation. Thus knowledge gained from large data pretraining can be effectively passed down to depression detection]

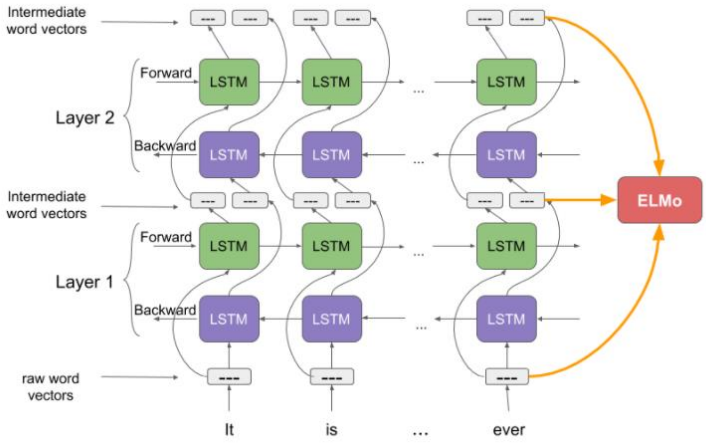
Mean, attention and max pooling -> depression detection

Background knowledge to understand more

Continuous Skip Grams



ELMo vector assigned to a token or word is actually a function of the entire sentence containing that word



fastText

Instead of learning vectors for words directly, fastText represents each word as an n-gram of characters. So, for example, take the word, "*artificial*" with n=3, the fastText representation of this word is <*ar, art, rti, tif, ifi, fic, ici, ial, al*>, where the angular brackets indicate the beginning and end of the word.

Patient Health Questionnaire score [PHQ-8]

Over the last 2 weeks, how often have you been bothered by any of the following problems? (Use "✓" to indicate your answer)	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down	0	1	2	3
7. Troubleconcentratingonthings,suchas reading thenewspaperorwatchingtelevision.	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

(For office coding: Total Score _____ = ___ + ___ + ___)

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MDPHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues. For research in formation, contact Dr. Spitzera trls8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc. Copyright© 1999 Pfizer Inc. All rights reserved. Reproduced with permission

[https://www.psychologywizard.net/uploads/2/6/6/4/26640833/kroenke_phq8.pdf]

5 fold cross validation

K-Fold CV is where a given data set is split into a K number of sections/folds where each fold is used as a testing set at some point.