Unveiling Mental Health Insights through Social Media Analysis

DATS 6312 – Natural Language Processing

Sunisha Harish

Contents

Introduction

In the year 2020, the World Health Organization (WHO) brought attention to the widespread impact of mental disorders, revealing that nearly 1 billion individuals, surpassing one in ten of the global population, grapple with the intricacies of mental health. Within the United States, the year 2021 unfolded with a notable statistic: over one in five adults, totaling a staggering 57.8 million individuals, reported living with a mental health condition. This metamorphosis not only positions social media platforms as vital hubs for connection but also as distinctive arenas for discerning and comprehending the nuanced indicators of mental health in our increasingly interconnected world. As our digital landscapes undergo continual evolution, the convergence of mental health and social media emerges as an increasingly vital frontier for exploration and analysis.

Dataset

Dataset Used

Upon formulating our problem statement, we embarked on a comprehensive literature review to investigate the various solutions that have been applied to address this challenge. During our exploration, we encountered a research paper available at (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10315126) that aligns closely with our specified requirements. Recognizing the compatibility of their data with our research objectives, we opted to leverage the data presented in this research paper as a valuable foundation for our own investigation.

The tweets were classified into 10 different classes based on the mental health disorder. The classes were: EATING DISORDER, SCHIZOPHRENIA, OCD, PTSD, ANXIETY, BIPOLAR, AUTISM, DEPRESSION, ADHD, CONTROL. Control Group was the non diagnosed group of users.

Splits

We did a standard split of training-test and dev. The training set was used for training the model and the test set was used to make decisions while training the model. The dev set was untouched and only evaluated on right before the project presentation.

My Contributions

Combining the Dataset

Due to the extensive size of the dataset, it was divided and provided to us in distinct folders, each corresponding to a specific disorder group. Within these folders were multiple .csv files, each containing tweets from individual users. Additionally, a separate partition .csv file was supplied, detailing the files utilized for training and testing, along with pertinent information such as file names and associated disorder classes. I wrote a python code to combine the dataset and generate train and test data files.

Train Bert on the English Dataset

I was responsible for training the Bert model on the English tweets dataset. I used the bert-base-uncased model and BertForSequenceClassfication to finetune based on our dataset.

I also tried bert-base-multilingual-uncased model to try incorporating and training both English and Spanish datasets but this was not giving me proper results.

```
def model_definition():

    model = BertForSequenceClassification.from_pretrained(MODEL_NAME_NLP, num_labels=10)  # 10 classes
    model = model.to(device)
    #optimizer = torch.optim.AdamW(model.parameters(), lr=LR)
    param_optimizer = list(model.named_parameters())
    no_decay = ['bias', 'LayerNorm.weight']
    optimizer_grouped_parameters = [
        {'params': [p for n, p in param_optimizer if not any(nd in n for nd in no_decay)],
         'weight_decay_rate': 0.1},
        {'params': [p for n, p in param_optimizer if any(nd in n for nd in no_decay)],
         'weight_decay_rate': 0.0}
    ]
    optimizer = AdamW(optimizer_grouped_parameters,
                      lr=2e-5,
                      eps=1e-8
                      )
    criterion = nn.CrossEntropyLoss()
    scheduler = ReduceLROnPlateau(optimizer, mode='min', factor=0.5, patience=LR_PATIENCE, verbose=True)

    print(model, file=open(f'summary_{MODEL_NAME}.txt', 'w'))

    return model, optimizer, criterion, scheduler
```

Figure 1: Model definition for English Tweets model

Test Bert Script for the model

I was responsible for creating the test script for the trained models so that our models could be evaluated on test and dev dataset.

Results

Model	Accuracy	F1Score
bert-base-uncased	0.65	0.62

The model did not perform very well because the Twitter data is not proctored and the context and structure of tweets for few disorder classes tend to be similar.

Conclusion

In conclusion, our project wasn't able to successfully classify English language and Spanish language tweets using the BERT based classification. One main reason is similar context in different classes.

Further Improvements

- The development of a mental health assistance chatbot to provide necessary support for the user proved to be challenging for us. However, we plan to try to incorporate this in the future.

- We can also try implementing the langdetect model to detect the language in which a tweet was posted and classify it accordingly.

- We could try langdetect model to detect the language in which the tweet was posted and then use the classification model accordingly.

REFERENCES

1. Bert-base-uncased · hugging face. bert-base-uncased · Hugging Face. (n.d.-a). https://huggingface.co/bert-base-uncased
2. M. E. Villa-Pérez, L. A. Trejo, M. B. Moin and E. Stroulia, "Extracting Mental Health Indicators from English and Spanish Social Media: A Machine Learning Approach," in IEEE Access, doi: 10.1109/ACCESS.2023.3332289.
3. Mental Illness Classification on social media texts using Deep Learning ... (n.d.). https://www.researchgate.net/publication/361757484_Mental_Illness_Classification_on_ Social_Media_Texts_using_Deep_Learning_and_Transfer_Learning

Percentage of code copied from the internet: Around 40% of code was copied from the internet.