

Depression Detection from Social Media

Kaishuo Wang, 101289185

1. Introduction

Depression is a common mental health issue that can have severe consequences if left untreated. Detecting signs of depression early from social media text could help identify at-risk individuals and connect them with support and treatment. This is the motivation behind the DepSign-LT-EDI@ACL-2022 shared task on detecting signs of depression from social media posts. The task involves classifying posts into three categories - "not depressed", "moderately depressed", and "severely depressed" - based on the textual content. Automated methods for this classification could assist in screening and identification of depressed individuals from large volumes of social media data. For this project, I'm planning to modify the method proposed by (Wang et al., 2022) from two aspects.

2. Related Work

(Wang et al., 2022) proposed a structure to utilize sentiment analysis and sentence embeddings to represent emotional and semantic aspects of the texts. Specifically, they generate sentiment features (i.g., VDA scores) using VADER while sentence embeddings were derived from pre-trained language models. These features were then fed into three methods including gradient boosting models, pre-trained models, and contrastive pre-trained models to classify the three depression levels, finally applying ensemble techniques for integrating these predictions. However, as discussed in (Saha et al., 2022), VADER has limitations in capturing nuanced sentiment compared to contextual models like BERT which have shown stronger performance on sentiment analysis tasks.

Moreover, (Deng et al., 2022) proposed a novel way to transform text classification problems into masked language modelling by constructing appropriate prompts. When applied to the SemEval-2022 Task 4, their prompt-based method showed promising results compared to other models.

3. Overview

To address the limitations of prior work, I propose two main modifications:

1. Using a fine-tuned BERT model rather than VADER to generate richer sentiment features. Specifically, I will fine-tune BERT on sentiment analysis datasets to adapt it to this task. The contextual embeddings from BERT are expected to better capture nuanced emotions compared to the lexicon-based VADER approach.
2. Exploring prompt-based classification following (Deng et al., 2022) rather than gradient boosting and contrastive pre-trained models. This involves carefully constructing prompts like "The person feels [MASK] based on this post:" where the model must fill in "not depressed", "moderately depressed" or "severely depressed". Fine-tuning BERT on these prompts could produce accurate classification without extensive task-specific architecture modifications.

I hypothesize that these two improvements will enhance the representation of the emotional content and allow the model to have a better performance.

3. Dataset

I will use the publicly available DepSign-LT-EDI@ACL-2022 dataset containing 8891 sentences for training, 4496 for validation, and 3245 for testing. The training and validation sets have 3801 "not depressed", 8325 "moderately depressed", and 1261 "severely depressed" samples. Each sample has the text and a label indicating the depression level. This dataset will allow comprehensive training and evaluation of the proposed methods.

Reference

1. Wang, W. Y., Tang, Y. C., Du, W. W., & Peng, W. C. (2022, May). NYCU_TWD@ LT-EDI-ACL2022: Ensemble models with VADER and contrastive learning for detecting signs of depression from social media. In *Proceedings of the second workshop on language technology for equality, diversity and inclusion* (pp. 136-139).
2. Saha, S., Showrov, M. I. H., Rahman, M. M., & Majumder, M. Z. H. (2022, September). VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak. In *International Conference on Machine Intelligence and Emerging Technologies* (pp. 371-385). Cham: Springer Nature Switzerland.
3. Deng, Y., Dou, C., Chen, L., Miao, D., Sun, X., Ma, B., & Li, X. (2022). BEIKE NLP at SemEval-2022 Task 4: Prompt-Based Paragraph Classification for Patronizing and Condescending Language Detection. *arXiv preprint arXiv:2208.01312*.