

NYCU_TWD@LT-EDI-ACL2022: Ensemble Models with VADER and Contrastive Learning for Detecting Signs of Depression from Social Media

Wei-Yao Wang^{†,1}, Yu-Chien Tang^{†,2}, Wei-Wei Du^{†,2}, Wen-Chih Peng¹

National Yang Ming Chiao Tung University, Hsinchu, Taiwan

¹{sf1638.cs05, wcpeng}@nctu.edu.tw, ²{tommytyc, wwdu}.cs10@nycu.edu.tw

Abstract

This paper presents a state-of-the-art solution to the LT-EDI-ACL 2022 Task 4: *Detecting Signs of Depression from Social Media Text*. The goal of this task is to detect the severity levels of depression of people from social media posts, where people often share their feelings on a daily basis. To detect the signs of depression, we propose a framework with pre-trained language models using rich information instead of training from scratch, gradient boosting and deep learning models for modeling various aspects, and supervised contrastive learning for the generalization ability. Moreover, ensemble techniques are also employed in consideration of the different advantages of each method. Experiments show that our framework achieves a 2nd prize ranking with a macro F1-score of 0.552, showing the effectiveness and robustness of our approach.

1 Introduction

Social media enable people to communicate and acquire information regardless of distance due to the rapid growth of the Internet. Besides, people can express their emotions about posts, news, and discussions on social media through texts and videos, which has thus attracted researchers who are interested in analyzing the emotional behavior of user comments. For instance, Saha et al. (2021) introduced a speech act classification Twitter dataset and presented an attention mechanism to incorporate intra-modal and inter-modal information. AudiBERT, which adopts the multimodal nature of the human voice, was proposed to screen depression (Toto et al., 2021)., and Pirayesh et al. (2021) proposed a social-contagion based framework based on meta-learning for early detection of depression.

In this challenge hosted by LT-EDI¹, given social media posts in English, the goal is to detect the

signs of depression and classify them into three labels, namely *not depression*, *moderate*, and *severe*. To tackle the shared task, we propose a framework with three methods for modeling the given texts. Specifically, the sentence embedding is produced by pre-trained models, and the VAD score (positive, neutral, negative, and compound) is generated by VADER (Hutto and Gilbert, 2014). Then, our first method utilized sentence embedding and VAD scores in gradient boosting models using SMOTE (Chawla et al., 2002) to mitigate the imbalance issue. The second method used a multi-layer perceptron (MLP) to fine-tune the pre-trained models. In addition, the third method further incorporated VAD embedding with MLP to classify the signs of depression. Furthermore, the third method adopted supervised contrastive learning (Gunel et al., 2021) in both sentence embedding and VAD embedding to enhance the capability of generalization. Afterwards, we used ensemble techniques, which have been used for substantially improving model performance (Wang et al., 2020; Wang and Peng, 2022), to consider the advantage of each method for boosting the performance.

We use the dataset provided by (Sampath et al., 2022) to detect the signs of depression from social media text. The dataset contains 8,891 posts for training, 4,496 posts for validation, and 3,245 posts for evaluation, while each sample is composed of three columns: *PID*, *Text*, and *Label*. Table 1 shows some examples of the dataset.

In summary, our main results and observations are described as follows:

- We propose a framework with three methods including gradient boosting models, fine-tuning pre-trained models, and fine-tuning pre-trained models by supervised contrastive learning for modeling different aspects.
- Besides, the VAD score provides additional

[†]Equal contributions.

¹<https://sites.google.com/view/lt-edi-2022/home>

Table 1: Samples from the depression dataset.

PID	Text	Label
train-pid-1	My life gets worse every year : That's what it feels like anyway...	moderate
train-pid-2	Words can't describe how bad I feel right now : I just want to fall asleep forever.	severe
train-pid-3	Is anybody else hoping the Coronavirus shuts everybody down?	not depression

sentiment scores for detecting the signs of depression, and we adopt ensemble techniques to take advantage of each model.

- Our ensemble method achieved competitive performance in the shared task and won the 2nd prize (0.552 macro F1-score) in detecting signs of depression from social media text.

2 Related Work

Social media are among the platforms used to express one’s emotions. They can therefore be viewed as an environment to study and discover user feelings. Recently, there have been several approaches to detecting signs of depression to eliminate the negative impact of emotions. For instance, [Toto et al. \(2021\)](#) introduced a framework with transfer learning to the multi-modality of textual context and audio characteristics of the human voice. [Zogan et al. \(2021\)](#) proposed DepressionNet by summarizing history posts as a summary of the user and applying different modalities to infer user behavior, which motivated us to include VAD scores as the additional post feature in this challenge.

3 Method

Figure 1 illustrates the pipeline of our framework. Given the input text, we first generate sentiment features (i.e., VAD scores) by VADER and sentence embeddings from pre-trained models. Then, we adopt three methods to model various aspects of the text, and apply ensemble techniques for integrating these predictions. Specifically, we use an unsupervised sentiment prediction, VADER, to assign sentiment scores to each sentence for measuring the sentiment effect of the word.

3.1 Method 1: Gradient Boosting Models

We use SentenceTransformers ([Reimers and Gurevych, 2019](#)) to generate pre-trained sentence

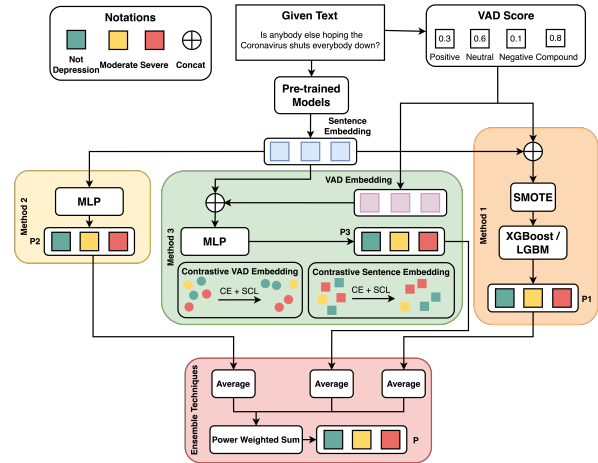


Figure 1: The illustration of our proposed framework.

embeddings, and concatenate the sentiment feature embeddings and pre-trained sentence embedding to take different perspectives into account. Besides, SMOTE ([Chawla et al., 2002](#)) and CondensedNearestNeighbour ([Gowda and Krishna, 1979](#)) is used for tackling the imbalanced classification problem. Then, LightGBM ([Ke et al., 2017](#)) and XGBoost ([Chen and Guestrin, 2016](#)) are applied as classifiers to predict the probability of each category in order to reduce the bias and variance through combining different learners. Cross-entropy is applied to optimize the values of the hyper-parameters.

3.2 Method 2: Pre-Trained Models

Fine-tuning pre-trained language models has demonstrated success in a wide range of natural language tasks since they provide fruitful information without the effort of training from scratch. To this end, we use three different pre-trained language models for fine-tuning in this task, including RoBERTa ([Liu et al., 2019](#)), ELECTRA ([Clark et al., 2020](#)), and DeBERTa ([He et al., 2021](#)). Specifically, for each pre-trained model, each given text is first tokenized and then produces the sentence embedding. Then, the sentence embedding is fed into a MLP to generate the predicted probabilities of depression. To tackle the imbalance issue, we employ torchsampler² for rebalancing the class distributions. The objective function is trained to minimize the cross-entropy, and the pre-trained models are applied from ([Wolf et al., 2019](#)).

²<https://github.com/ufoyim/imbalanced-dataset-sampler>

3.3 Method 3: Contrastive Pre-Trained Models

To combine the ideas of the previous two methods, the sentence embedding is generated in the same way as the Sec. 3.2. Moreover, we apply VAD scores through an embedding layer with GeLU activation function (Hendrycks and Gimpel, 2016), which has been used in several natural language tasks. Afterwards, we concatenate the sentence embedding and VAD embedding as the input of an MLP to classify the probabilities of each sign of depression. The imbalance technique is also used as in Sec. 3.2.

We jointly train supervised contrastive learning (Gunel et al., 2021) and cross-entropy for enhancing the generalization of our method. Specifically, sentence embeddings and VAD embeddings are adopted supervised contrastive learning, respectively. Thus, similar sentences would become closer, while irrelevant sentences would increase the distance. The VAD scores would follow this phenomenon since it is reasonable that similar sentiment features would have a closer distance compared to the dissimilar sentiment features.

3.4 Ensemble Techniques

To combine the different advantages of each model, soft-voting ensemble is used for ensembling each method. Specifically, the predicted probabilities of Method 1 P_1 are averaged by LightGBM and XGBoost, and the predicted probabilities of Method 2 P_2 are averaged by RoBERTa, ELECTRA, and DeBERTa. The predicted probabilities of Method 3 P_3 are weighted averaged by RoBERTa, ELECTRA, and DeBERTa with the weights of 0.15, 0.5, and 0.35, respectively.

To boost the performance, the final predicted probabilities P are computed with power weighted sum as in (Wang and Peng, 2022):

$$P = P_1^N \times w_1 + P_2^N \times w_2 + P_3^N \times w_3, \quad (1)$$

where w_1, w_2, w_3 are weights of the corresponding model, and N is the weight of power. In this paper, we tune these hyper-parameters based on the validation set and use N as 4 and ensemble weights as 1.00, 0.67, and 0.69, respectively.

4 Experiments

4.1 Implementation Details

Due to the page limit, we report the selected hyper-parameters of each method and the official code in

the appendix³. It is noted that all hyper-parameters are tuned with the validation set by grid search.

4.2 Depression Performance

We first examine the advantages of each model, and Table 2 reports the F1-score of each category of each method in the validation set. It is observed that each model specializes in detecting various signs of depression, respectively. For instance, gradient boosting models are adept at identifying *not depression*. As a result, ensemble techniques incorporate different models to improve performance and robustness.

The results for the testing set are shown as Table 3 in terms of accuracy and macro-F1. Our ensemble model performs the best compared to each method we introduced and won 2nd prize among all the participants.

5 Conclusion

In this paper, we introduce a framework for the detecting signs of depression from social media text challenge which incorporates three different methods, namely gradient boosting models, pre-trained models, and contrastive pre-trained models. Furthermore, ensemble techniques are adopted to enable our model’s ability to integrate the strengths of each model. The experimental results demonstrate the effectiveness of our framework and verify the different capabilities of each method. Thus, ensembling three approaches achieves better performance on both the validation set and the testing set, resulting in a second ranking and achieving a competitive performance.

References

- Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16:321–357.
- Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *KDD*, pages 785–794. ACM.
- Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: pre-training text encoders as discriminators rather than generators. In *ICLR*. OpenReview.net.
- K. Chidananda Gowda and G. Krishna. 1979. The condensed nearest neighbor rule using the concept of

³<https://github.com/wywyWang/Depression-Detection-LT-EDI-ACL-2022>

Table 2: F1-score of each category of each method for the validation set.

	Gradient boosting models	Pre-trained models	Contrastive pre-trained models	Ensemble model
Not Depression	0.638	0.578	0.613	0.630
Moderate	0.633	0.704	0.667	0.707
Severe	0.416	0.510	0.506	0.532
Macro-F1	0.562	0.597	0.595	0.623

Table 3: Performance of our approach for the testing set.

	Gradient boosting models	Pre-trained models	Contrastive pre-trained models	Ensemble model
Accuracy	0.571	0.635	0.597	0.633
Macro-F1	0.496	0.528	0.523	0.552

- mutual nearest neighborhood (corresp.). *IEEE Trans. Inf. Theory*, 25(4):488–490.
- Beliz Gunel, Jingfei Du, Alexis Conneau, and Veselin Stoyanov. 2021. Supervised contrastive learning for pre-trained language model fine-tuning. In *ICLR*. OpenReview.net.
- Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. DeBERTa: decoding-enhanced bert with disentangled attention. In *ICLR*. OpenReview.net.
- Dan Hendrycks and Kevin Gimpel. 2016. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415.
- Clayton J. Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*. The AAAI Press.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *NIPS*, pages 3146–3154.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.
- Jahandad Pirayesh, Haiquan Chen, Xiao Qin, Wei-Shinn Ku, and Da Yan. 2021. Mentalspot: Effective early screening for depression based on social contagion. In *CIKM*, pages 1437–1446. ACM.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *EMNLP/IJCNLP (1)*, pages 3980–3990. Association for Computational Linguistics.
- Tulika Saha, Apoorva Upadhyaya, Sriparna Saha, and Pushpak Bhattacharyya. 2021. Towards sentiment and emotion aided multi-modal speech act classification in twitter. In *NAACL-HLT*, pages 5727–5737. Association for Computational Linguistics.
- Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022. Findings of the shared task on Detecting Signs of Depression from Social Media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association Computational Linguistics.
- Ermal Toto, M. L. Tlachac, and Elke A. Rundensteiner. 2021. Audibert: A deep transfer learning multimodal classification framework for depression screening. In *CIKM*, pages 4145–4154. ACM.
- Wei-Yao Wang, Kai-Shiang Chang, and Yu-Chien Tang. 2020. Emotiongif-yankee: A sentiment classifier with robust model based ensemble methods. *CoRR*, abs/2007.02259.
- Wei-Yao Wang and Wen-Chih Peng. 2022. Team yao at factify 2022: Utilizing pre-trained models and co-attention networks for multi-modal fact verification. *CoRR*, abs/2201.11664.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *CoRR*, abs/1910.03771.
- Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guangdong Xu. 2021. Depressionnet: Learning multimodalities with user post summarization for depression detection on social media. In *SIGIR*, pages 133–142. ACM.