2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385946

An Early Depression Detection Model on Social Media using Emotional and Causal Features

Linlin Zong[1], Jian Zheng[1], Xianchao Zhang[1], Xinyue Liu[1], Wenxin Liang[1], Bo Xu[2*]
[1]*School of Software, Dalian University of Technology, Dalian, China*
[2]*School of Computer Science and Technology, Dalian University of Technology, Dalian, China*
xubo@dlut.edu.cn

Abstract—**Early depression detection is essential to enable healthcare professionals to effectively intervene and treat the depressive conditions. In existing research, an increasing number of psychological manifestations of depression are incorporated, and demonstrated effective by integrating them into computational models, particularly the emotional cues of depression. Although emotional factors are effective in depression detection, few studies have paid attention to the underlying depressive causal factors hidden within social media text for depression detection. In this paper, we propose a novel partitioned filtering network model to extract emotional and causal features for predicting the depressive users on social media. Experimental results demonstrate the proposed model achieves superior performance over recent baseline models on the dataset by highlighting the emotional and causal factors.**

Index Terms—**Early Depression Detection, Natural Language Processing, Social media, Deep Learning**

I. Introduction

Early depression detection and timely intervention can assist patients more effectively coping with depressive conditions and alleviating somatic symptoms [1]. Traditionally, early diagnosis of depression relied on clinical judgments by medical professionals and the use of assessment scales, such as the Hamilton Depression Rating Scale (HAMD) [2], the Beck Depression Inventory (BDI) and the Center for Epidemiologic Studies Depression Scale (CES-D). However, these methods possess inherent subjectivity, potentially leading to delays in providing timely treatments [3]. With the rising popularity of social media, people are more inclined to share their lives, opinions, and emotions online, which provides significant potential for exploring early depression detection features. To this end, researchers have turned to natural language processing (NLP) techniques on social media to develop more accurate and applicable methods for depression detection.

Early psychological research has found that depressed persons tend to exhibit distinct behaviors and speech patterns compared to non-depressed ones [4]–[6]. Based on this finding, recent studies have shown that depressed social media users tend to frequently use absolutist words and negative emotional vocabulary [7] , such as 'hate' and 'suicide', and be more inclined to use the pronoun 'I' [8]. These language patterns further influence depressed patients' living habits. For example, they tend to express negative emotions, resulting in releasing social media text that exhibits significant emotional cues of depression. To capture depressive emotions on social media, the emotional characteristics of individual depressed users are modeled, such as the ratio of positive and negative emotions and the temporal changes of emotions over a period [8], [9]. Given the established correlation between emotional expression and psychological states [10], [11], numerous studies have begun investigating the incorporation of emotional features into automatic depression detection [10], [11] and yielded noteworthy achievements.

Beyond emotional features, causal factors of depressed emotions are determining and valuable on detecting depressed users and intervening mental disorders. Depressed social media users often expressed the causes of their emotional conditions when describing their affliction. The causal features can be thus used to depict the emotional evolution law of depressed events, and then help to trace the causal development of depression, which contributes to better generating clinical advice and intervening solutions. Despite the prevalence of research focused on exploring the role of emotional features in depression detection, causal factors of emotions are unexplored in prior studies.

To model the causal factors, we propose an emotion-cause mutual interacted model for depression detection. Our model first employs BERT to obtain semantic representations of posts, then input the semantic representations into a partitioned filtering network to extract emotional and causal features, and finally combines semantic features and emotional causal features for depression detection. Experimental results demonstrate that our model outperforms other baseline models in terms of experimental performance.

II. Methodology

Given a user C, we obtain his or her historical tweets in chronological order. We represent all the tweets of C with $P = \{p_1, p_2,\dots, p_N\}$, where N represents the total number of tweets. In this paper, our aim is to identify individuals with depression by analyzing the tweets and incorporating the semantic features $F^s = \{f^s_1, f^s_2,\dots, f^s_N\}$, emotional features $F^e = \{f^e_1, f^e_2,\dots, f^e_N\}$, and the proposed causal features $F^c = \{f^c_1, f^c_2,\dots, f^c_N\}$ from each post within the tweets. Therefore, the task of early depression detection can be formulated as a binary classification task on predicting the depression tendency label $\hat{y}$ of the user C. The overall architecture of the Emotion Cause Detection model (ECD) is illustrated in Fig. 1.

979-8-3503-3748-8/23/$31.00 ©2023 IEEE

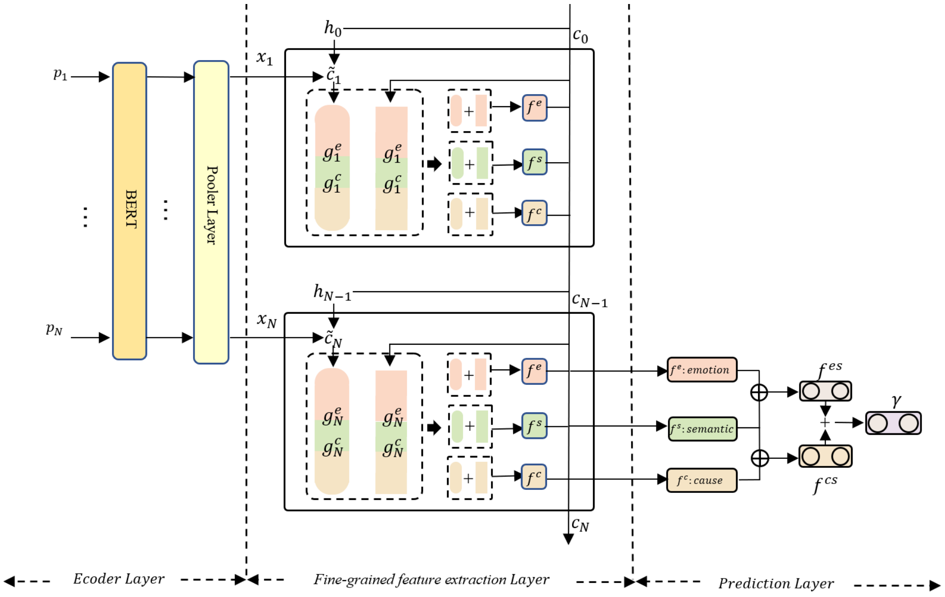

Fig. 1. Architecture of ECD. ECD consists of three components: the encoder layer, fine-grained feature extraction layer, and prediction layer.

A. Encoder Layer

In terms of text feature representation, we leverage the pretrained BERT language model as the underlying encoder to yield contextualized clause representations. For the user s_j, which corresponds to a sample in our dataset, we obtain all the post representations $X = \{x_1^{s_j}, x_2^{s_j}, \ldots, x_N^{s_j}\}$.

B. Fine-grained Feature Extraction

The network structure is a recurrent feature encoder. At the i-th time step, each unit of the network receives the historical cell state c_{i-1} and hidden state h_{i-1} passed from the previous $(i-1)$-th time step's unit. And we can obtain the representation of the current unit's candidate information $\tilde{c}_i$ using $\tilde{c}_i = \tanh(\text{Linear}([x_i; h_{i-1}]))$. To extract fine-grained features, we utilize two feature-related gates $g_i^e = \text{Cummax}(\text{Linear}([x_i; h_{i-1}]))$ and $g_i^c = 1 - \text{Cummax}(\text{Linear}([x_i; h_{i-1}]))$ to partition the feature, where $\text{Cummax}(.) = \text{Cumsum}(\text{Softmax}(.\ ))$.

Both the current and historical information undergo feature partitioning through their corresponding gates. The current information $\tilde{c}_i$ will be divided into three partitions: current emotion-related partition $\rho_{e,\tilde{c}_i} = g_i^e \circ g_i^c$, current cause-related partition $\rho_{c,\tilde{c}_i} = g_i^e - \rho_{s,\tilde{c}_i}$, and current semantic-related partition $\rho_{s,\tilde{c}_i} = g_i^c - \rho_{s,\tilde{c}_i}$, where $\circ$ denotes element-wise multiplication. The historical emotion-related partition $\rho_{e,c_{i-1}}$, historical cause-related partition $\rho_{c,c_{i-1}}$, and historical semantic-related partition $\rho_{s,c_{i-1}}$ for historical information cell c_{i-1} are also generated in the same manner.

Next, we can obtain three feature representations combining current information cell $\tilde{c}_i$ and historical information cell c_{i-1}: emotional feature $f_i^e = \tanh(\rho_{e,c_{i-1}} \circ c_{i-1} + \rho_{e,\tilde{c}_i} \circ \tilde{c}_i)$, causal feature $f_i^c = \tanh(\rho_{c,c_{i-1}} \circ c_{i-1} + \rho_{c,\tilde{c}_i} \circ \tilde{c}_i)$ and semantic feature $f_i^s = \tanh(\rho_{s,c_{i-1}} \circ c_{i-1} + \rho_{s,\tilde{c}_i} \circ \tilde{c}_i)$.

Furthermore, the current time step's output, the hidden state $h_i = \tanh(c_i)$ and the cell state $c_i = \text{Linear}([f_i^e, f_i^c, f_i^s])$, is derived from these three partitions.

C. Prediction Layer

We concatenate the emotional features with the semantic features to derive the emotion-specific features f^{es}, and then we concatenate the causal features with the semantic features to derive the cause-specific features f^{cs}. Then the final representation is $\gamma = f^{es} + f^{ec}$. We feed γ into a fully connected layer (FC) to obtain the final prediction $\hat{y} = \sigma(\text{FC}(\gamma))$. The loss function of the depression detection task can be formulated as:

$$L = -\sum_{i=1}^{N}(y \log(\hat{y}) + (1-y)\log(1-\hat{y})), \tag{1}$$

where N denotes the number of social media users.

III. Experiments

A. Experimental Settings

We conduct the experiment on the eRisk2018 shared task one dataset. We performed data cleaning on each post, breaking the contents into individual sentences, and manually removing irrelevant or disruptive contents, such as non-English text, URL links. We retain emojis interspersed in the text in a separate field *Emoji*, considering that the use of emojis may indicate stereotyped behaviors of depression users.

In order to validate the effectiveness of our model, we compared with a series of baseline approaches as follows: BiLSTM [12], BERT [13], BioBERT [14], MentalBERT [15], EAN [16], ERAN [17]. To ensure fair comparisons, we maintained the same hyperparameter settings of each baseline model. Specifically, we set the maximum number of training epochs to be 30, the batch size to be 4, the dropout rate to be 0.1, and the hidden layer size of the feature extraction to be 200. Moreover, we employed a learning rate of 1e-5 along with the linear schedule witch warm-up.

TABLE I
EXPERIMENT RESULT

Method	P	R	F1
BiLSTM	0.50	0.13	0.21
BERT	0.64	**0.47**	0.54
BioBERT	0.58	**0.47**	0.52
MentalBERT	0.56	0.33	0.42
EAN	0.50	0.23	0.32
ERAN	0.62	0.38	0.47
ours(ECD)	**0.78**	**0.47**	**0.59**

B. Results and analysis

The precision(P), recall(R), and F1 score(F1) are presented in Table I, where the best results for each metric are highlighted in bold. The experimental results demonstrate that our model outperforms other baseline models on the dataset, validating the effectiveness of our concept of effectively partitioning emotion and cause features in text to assist in depression pattern detection.

IV. CONCLUSION

With the development of social media platforms, more and more people tend to share posts containing personal feelings. These texts contain potential patterns that can differentiate between individuals with normal emotions and depressed emotions. In this work, we introduce the causal features into a computation model ECD to detect early depression. Based on our experiments, our model outperforms other baseline models, achieving higher precision, recall, and F1 scores. Future studies can be extended by extracting more fine-grained emotional features and generating useful clinical advice for social media users.

ACKNOWLEDGMENT

This work was supported by the Social Science Planning Foundation of Liaoning Province under Grant L21CXW003.

REFERENCES

[1] Centers for Disease Control, Prevention (CDC, et al. Current depression among adults—united states, 2006 and 2008. *MMWR. Morbidity and mortality weekly report*, 59(38):1229–1235, 2010.

[2] Max Hamilton. A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, 23(1):56, 1960.

[3] Xiuzhuang Zhou, Kai Jin, Yuanyuan Shang, and Guodong Guo. Visually interpretable representation learning for depression recognition from facial images. *IEEE transactions on affective computing*, 11(3):542–552, 2018.

[4] Jesper Pedersen, JTM Schelde, E Hannibal, K Behnke, BM Nielsen, and M Hertz. An ethological description of depression. *Acta psychiatrica scandinavica*, 78(3):320–330, 1988.

[5] Luciano Fossi, C Faravelli, and M Paoli. The ethological approach to the assessment of depressive disorders. *The Journal of nervous and mental disease*, 172(6):332–341, 1984.

[6] Peter Waxer. Nonverbal cues for depression. *Journal of Abnormal Psychology*, 83(3):319, 1974.

[7] Mohammed Al-Mosaiwi and Tom Johnstone. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4):529–542, 2018.

[8] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133, 2004.

[9] Marc L Molendijk, Lotte Bamelis, Arnold AP van Emmerik, Arnoud Arntz, Rimke Haringsma, and Philip Spinhoven. Word use of outpatients with a personality disorder and concurrent or previous major depressive disorder. *Behaviour Research and Therapy*, 48(1):44–51, 2010.

[10] Nirmal Varghese Babu and E Grace Mary Kanaga. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. *SN Computer Science*, 3:1–20, 2022.

[11] Tianlin Zhang, Kailai Yang, Shaoxiong Ji, and Sophia Ananiadou. Emotion fusion for mental illness detection from social media: A survey. *Information Fusion*, 92:231–246, 2023.

[12] Alex Graves and Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[14] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[15] Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. Mentalbert: Publicly available pretrained language models for mental healthcare. *arXiv preprint arXiv:2110.15621*, 2021.

[16] Lu Ren, Hongfei Lin, Bo Xu, Shaowu Zhang, Liang Yang, and Shichang Sun. Depression detection on reddit with an emotion-based attention network: Algorithm development and validation. *JMIR Medical Informatics*, 9(7):e28754, 2021.

[17] Bin Cui, Jian Wang, Hongfei Lin, Yijia Zhang, Liang Yang, and Bo Xu. Emotion-based reinforcement attention network for depression detection on social media: Algorithm development and validation. *JMIR Medical Informatics*, 10(8):e37818, 2022.