Journal of Computers Vol. 33 No. 4, August 2022, pp. 95-106
doi: 10.53106/199115992022083304008

Depression Detection in Social Media using XLNet with Topic Distributions

Wang Gao[1*], Baoping Yang[2], Yuwei Wang[1], Yuan Fang[3]

[1] School of Artificial Intelligence, Jianghan University, Wuhan 430056, China
{gaowang2000, wangyuwei}@foxmail.com
[2] Physics and Telecommunications College of Engineering, Huanggang Normal University, Huanggang 438000, China
yangbp@hgnu.edu.cn
[3] School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China
fangyuan2000@foxmail.com

Received 19 September 2021; Revised 22 February2022; Accepted 3 March 2022

Abstract. Due to the complexity of depressive diseases, detecting depressed users on social media platforms is a challenging task. In recent years, with an increasing number of users of social media sites, this field of research has begun to develop rapidly. To improve the detection performance of traditional methods, two challenges need to be overcome. The first challenge is that textual content posted on social media platforms suffers from serious data sparseness. The second one is how to effectively use emotions, user information, and behavior characteristics to predict potentially depressed users. In this paper, we propose a novel model called the Topic-enriched Depression Detection Model (TDDM), which combines topic information and user behavior to predict depressed users on social media platforms. TDDM first employs a Conditional Random Field Regularized Topic Model (CRFTM) to extract the topic knowledge of user posts. XLNet is used to encode posts to further expand the semantic features of short texts. Finally, we integrate user behavior features into TDDM to improve the detection performance of the model. The experimental results on a real-world Twitter dataset demonstrate that the proposed model performs better than baseline models in detecting depressed users at both pseudo-document level and user level.

Keywords: depression detection, XLNet, topic model, BiLSTM

1 Introduction

Mental illness has become a serious problem that people all over the world will encounter. Statistics indicate that more than 50% of individuals will suffer from a mental illness at some point in their lives [1]. It is estimated that 20.6% of Americans aged 18 years and older (about one-fifth of adults) are diagnosed with mental illness or disorder in a certain year[1]. In addition, research results illustrated that the proportion of mass shooters with mental illness is significantly higher than the general population, which indicates that there is a relationship between mental illness and mass shootings [2]. As a more serious mental health problem than other mental illnesses, depression is a leading cause of disability worldwide [3]. According to a fact sheet provided by the World Health Organization, the risk of suicide in patients with depression is 20 times higher than non-depressed people[2]. Since the diagnosis of depression requires experienced psychiatrists to conduct comprehensive and detailed psychological tests, depression detection is usually a daunting challenge. Furthermore, in the early stages of depression, depressed people often do not seek help from psychiatrists.

Social media platforms allow users to express their feelings, opinions and thoughts in their own way. As a result, social media data has gradually become an effective tool for monitoring various public health issues such as depression. People with depression usually implicitly disclose their daily struggles due to depressive symptoms and their feelings of relieving stress on social media [4]. However, it is very time-consuming to detect individuals with depressive tendencies by manually screening their profiles and social media posts. Depression detection algorithms can automatically detect a large number of depressed people on social media, which helps prevent major fatalities that may occur in the future and provide them with medical assistance.

Since social media plays an important role in evaluating mental health, it has received great attention from many researchers in Natural Language Processing (NLP). These studies employ User-Generated Content (UGC) on social media such as Twitter or Facebook to detect depressed users, and achieve robust performance. For in-

1 https://nami.org/mhstats/
2 https://www.who.int/en/news-room/fact-sheets/detail/depression/

* Corresponding Author

stance, Gui et al. proposed a new cooperative multi-agent model to automatically select depression related data from a user's historical posts [5]. Chiu et al. first built a depression dictionary to collect UGC of depressed and non-depressed users on Instagram [6]. Then, a multi-modal system was constructed to identify users with depression tendencies by using textual, visual and behavior features.

Although previous studies have made great progress, they focused on modeling UGC along the timeline, ignoring the global topic knowledge in the posts. The topics that depressed users and non-depressed users pay attention to may be significantly different. Posts by individuals with depressive tendencies tend to discuss topics related to negativity, anxiety and emotional stress [7]. Furthermore, from a psychological point of view, individuals who have been indulged in negative emotions for a long time (usually more than one week), and are unable to carry out daily activities are most likely to experience depression symptoms such as insomnia, irritability, feelings of hopelessness and helplessness. Therefore, it is meaningful to consider topic knowledge representing global semantics in depression detection.

More importantly, in addition to UGC, previous psychological studies have shown that personality traits such as openness, responsibility and extroversion are associated with depression [8]. Similarly, personal sentiment and emotions can also indicate whether they are prone to depression. Park et al. found that users who are unemployed or low-educated are more likely to suffer from depression [4]. Previous studies either chose features that are not highly relevant or ignored significant features. For instance, Shen et al. captured several depression-related features, including not only user behaviors on social media, but also clinical diagnostic criteria for depression [3]. Based on dictionary learning, they proposed a novel dictionary learning model to learn user profile representations. Nevertheless, given the sparseness, high dimensionality and ambiguity of short texts in social media sites, dictionary learning is difficult to accurately extract the semantics of these short texts.

To tackle the above challenges, we propose a Topic-enriched Depression Detection Model (TDDM), which combines global topic knowledge and user behaviors for depression detection in social media. Since the information contained in a short text is not enough to determine whether a user is suffering from depression, we first concatenate the tweets of the same user to generate long pseudo-documents, and annotate these pseudo-documents based on the user's label. Secondly, a Conditional Random Field regularized Topic Model (CRFTM) [9] is employed to learn the topic knowledge of posts. CRFTM integrates semantic association into the process of topic inference to improve the probability of semantically correlated words appearing in the same topic. After that, the proposed model concatenates the last encoder layer hidden representation of XLNet [10] with the topic knowledge learned by CRFTM to form richer textual features. Finally, TDDM maintains the chronological order of pseudo-documents, and feeds them and user behavior features into a Bidirectional Long Short-Term Memory (BiLSTM) layer to detect depressed users in social media. To evaluate the pseudo-document level and user level performance of TDDM, we conduct extensive experiments on a labeled Twitter dataset. Experimental results demonstrate that our model achieves better performance than the state-of-the-art models, which shows the potential of TDDM to detect depression for more countries with different languages. In summary, our main contributions are as follows:

- We propose a novel Topic-enriched Depression Detection Model (TDDM) to detect depressed users in social media. TDDM first leverages CRFTM to capture the global topic knowledge of posts. Then, the topic information is incorporated into XLNet for pseudo-document level classification and user level classification. To the best of our knowledge, we are the first to integrate the topic knowledge based on CRFTM into XLNet.
- TDDM feeds post-level representations into BiLSTM to identify depressed users, which effectively utilizes information in the temporal dimension. Features extracted from user behaviors are also input into BiLSTM to capture the unique representation and long-term dependencies of the feature matrix.
- The performance of TDDM is compared with state-of-the-art baseline methods on a real-world Twitter dataset. Experimental results show that our model achieves the best precision, recall and F1-measure at both pseudo-document level and the user level.

2 Related Work

In this section, we will discuss the research related to the TDDM model and analyze the differences between them. Choudhury et al. pioneered the use of machine learning algorithms to detect Twitter users suffering from depression [20]. They analyzed the correlation between depressive content in social media and user behavior characteristics, and used it to classify potentially depressed users. Since then, relevant research has mainly focused on methods based on textual content features or user behavior features.

For methods based on textual content features, researchers extract various text features of user posts through statistical methods. Hu et al. proposed a new method for identifying depression in a large number of people [21]. They extracted behavior and text features from social media platforms to build machine learning models, and compared the accuracy of the models on different windows. Resnik et al. utilized topic models to extract text features and improved the performance of multiple methods on the task of detecting depression [7]. Their experimental results show that probabilistic topic models such as LDA can discover meaningful latent structures from document collections. As a result, topic models can improve the performance of depressed user detection tasks on social media platforms, and they found that more advanced topic models improve the task even more. Ríssola et al. analyzed a variety of factors that can be used to classify users with mental disorders and ordinary users on social media platforms [22]. Their experimental results show that there are significant differences in expressions and emotions between ordinary users and users with mental disorders on social media platforms.

For methods based on user behavior features, unlike tweet-level features captured from a single tweet, user-level features are captured from multiple posts of the same user. Shen et al. proposed a dictionary learning model to distinguish between depressed and non-depressed users in social media [3]. The model not only uses textual features, but also involves user behavioral features. Luan et al. analyzed the differences in cross-domain characteristics of depressed users on social media sites [23]. In addition, they proposed a cross-domain depression detection model, which can transmit relevant features across domains using adaptive strategies. Wu et al. hired a large number of heavy users of the social media site Facebook and collected various data about them [24]. Afterward, they analyzed the user characteristics, behavior characteristics, and textual characteristics of these accounts, and used them to detect depressed users on Facebook.

There are many studies using deep learning techniques to predict depressed users on social media platforms. Hussain et al. designed a framework incorporating deep learning to determine that social media information is an important factor in predicting depression [25]. Based on the framework, they proposed a detection model that applies deep learning technology to predict whether Facebook users are prone to depression. Lam et al. proposed a model based on topic modeling, and use Convolutional Neural Network (CNN) and transformer to extract acoustic feature information [26]. The experimental results show that their model can effectively extract text and audio features, and improve the performance of the depression recognition system. Orabi et al. proposed a classification model based on word vectors, which detects depressed users based on the textual content of social media [27]. Furthermore, they compared the performance of commonly used deep learning models on the task of depression recognition.

However, the above methods are difficult to solve the sparsity of short texts on social media platforms, resulting in poor depression detection performance. As a result, the proposed model utilizes a topic model CRFTM to mine the topic background knowledge of tweets to expand the semantics of short texts. To the best of our current knowledge, this is the first attempt to integrate CRFTM-based topic knowledge into the detection of depressed users on social media.

3 Methodology

Our method divides the depression detection process in social media into three phases. In the first phase, the proposed method trains a topic model on the entire corpus to infer the topic distribution of each short text. We design a method to combine XLNet with global topic knowledge in the second phase. Finally, TDDM inputs textual representations and user level features into a BiLSTM model to identify depressed users on social media platforms.

3.1 Topical Information Mining

Automatic extraction of topical knowledge from a large number of short text collections has been widely used in NLP tasks such as context analysis, user interest analysis and text classification [31-32]. Traditional topic extraction approaches, such as Latent Dirichlet Allocation (LDA) [28] and Dynamic Topic Model (DTM) [29], can automatically mine latent topical structure from lengthy documents like scientific papers or news articles. These topic modeling methods utilize word co-occurrence patterns to reveal the probability distributions of topic-word and document-topic. Nevertheless, the sparsity problem of short texts on social media platforms hinders the topic knowledge mining process of these models. To solve this problem, we leverage a topic model called CRFTM to discover the latent topic structure from short texts. The CRFTM model first merges short texts to build longer pseudo-documents, and then semantic associations between words are used to improve the coherence of topics.

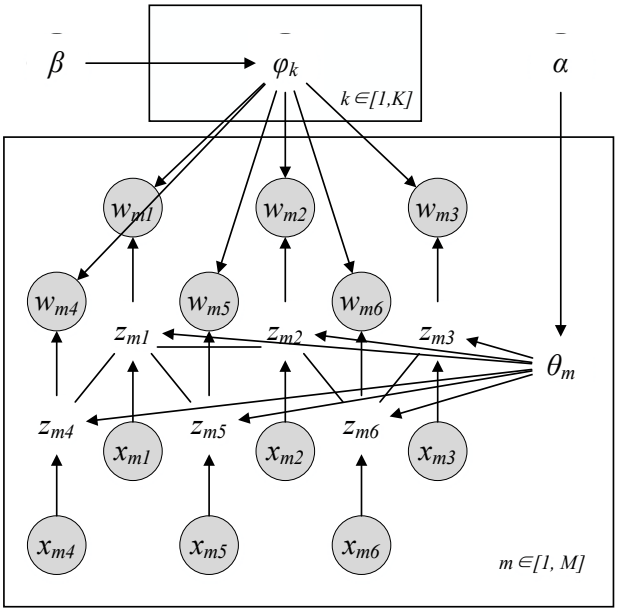

Fig. 1. The graphical representation of the CRFTM model

Specifically, the distance between different short texts is calculated by an embedding-based distance metric. The distance measure can identify semantically correlated terms (e.g., sad and unhappy) in two short texts [9]. If the two short texts are close to each other, they may belong to the same hidden topic. The k-medoid algorithm is then used to concatenate short texts into standard-length pseudo-documents. To further alleviate the sparsity, a constraint is added to the clustering algorithm, that is, the number of short texts aggregated for each pseudo-document is almost the same. After that, as shown in Fig. 1, the CRFTM model draws a document-specific topic distribution $\theta \sim Dir(\alpha)$ for pseudo-documents, and draws a term distribution $\phi_k \sim Dir(\beta)$ for each topic k. α and β are hyperparameters, which represent document-topic density and topic-word density, respectively. Finally, for each term index $n = 1, 2, ...N$ in pseudo-documents j, CRFTM draws a term $t_{j,n} \sim Mult(\phi_{z_{j,n}})$, and draws a topic label $\mathbf{z}_j$ as follow:

$$p(\mathbf{z}_j | \ \theta_j, \mathbf{r}_j) = \prod_{n=1}^{N} p(z_{j,n} | \ \theta) \Psi(z_{j,n}, r_{j,n}), \tag{1}$$

where the context-related words of $t_{j,n}$ are represented by $r_{j,n}$, and Ψ denotes the potential function used to promote semantically related words.

In the process of topic inference, the model utilizes collapsed Gibbs sampling to sample the topic assignment z by integrating out uncorrelated parameters:

$$p(z_{j,n} = k \mid \mathbf{z}_{j,-jn}, \mathbf{t}) \propto (c_{j,-jn}^{(t_{j,n})} + \alpha) \frac{c_{k,-jn}^{(t_{j,n})} + \beta}{c_{k,-jn} + V\beta} \Psi(z_{j,n} = k, r_{j,n}), \tag{2}$$

where $c_{\cdot,-jn}^{(\cdot)}$ denotes the count of term $t_{j,n}$ belonging to topic k, when term is removed from pseudo-document j or topic k, and V represents the dictionary size.

After enough sampling iterations, the model parameters θ and ϕ are estimated as follows:

$$\theta_{j,k} = \frac{c_j^{(k)} + \alpha}{\sum_{k=1}^{K} c_j^{(k)} + K\alpha}$$
$$\phi_{k,t} = \frac{c_k^{(t)} + \beta}{\sum_{t=1}^{V} c_k^{(t)} + V\beta} \quad , \tag{3}$$

where K represents the total number of topics. After obtaining the topic background knowledge, TDDM integrates it into the XLNet model.

3.2 Topic Knowledge Injection

This section describes in detail how the proposed method injects topic knowledge into XLNet. The proposed architecture of the depression detection task over social media streams is shown in Fig. 2. TDDM exploits the topic distribution captured from CRFTM and contextualized embeddings to learn effective representations for depression detection. The contextualized embeddings provided by XLNet can be used to solve the polysemy phenomenon where a word has multiple meanings. Traditional methods such as glove [11] and word2vec [12] leverage a fixed word vector to represent each word, and ignore their context. Therefore, they cannot address the ambiguity problem of polysemous words.

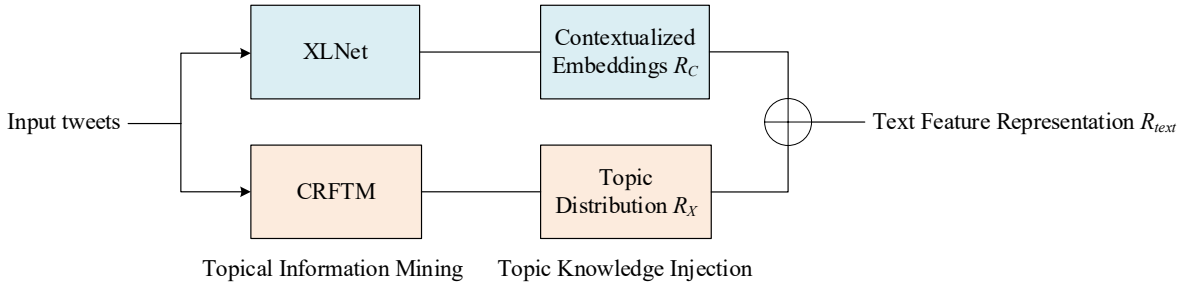

Fig. 2. Topic information mining and injection process in TDDM

Recently, transformer-based pre-training language models have achieved success in various fields. The transformer model utilizes a self-attention mechanism to perform attention calculations on the current word and other words in the sentence, which in turn obtains a better representation of the word. The mechanism learns the word embedding of the current word through its different attention to the context information. BERT proposed by Devlin et al. [13] is based on the transformer structure, and the contextualized representations it generates are applied to many NLP tasks such as text classification [14]. Nevertheless, in the training of BERT, the masked token prediction procedure is executed in parallel, which may lead to erroneous predictions. This is because the values of masked tokens may be dependent on each other, and parallel prediction cannot take advantage of the dependency. For example, given the message "She wrote a [MASK] with a [MASK]", using "letter" and "pencil" to fill in the masks is more likely than using "letter" and "bicycle". When BERT predicts the second mask, it ignores the first masked token, which may lead to an incorrect prediction result.

To solve the problem of BERT, XLNet utilizes the transformer-XL [15] architecture and incorporates permutation language modeling, permuting and combining each token to generate all possible word orders. Based on the transformer architecture, transformer-XL applies a new positional encoding and a segment-level recurrence method. During the training process, Transformer-XL first stores the hidden state produced by the previous segment. Then, when calculating the next new segment, the cached hidden state is used as the context. Similar to traditional language models, the permutation language modeling mechanism predicts the next token through a given context in the pre-training stage. The difference from BERT is that the mechanism does not predict tokens sequentially, but predicts them in random order. Therefore, the artificial symbol [MASK] is not necessary for the permutation language modeling, and it will not lead to inconsistent pretrain-finetuning.

In the TDDM model, XLNet is used to obtain contextualized embeddings R_X of the input short text. After

training on the entire corpus, CRFTM can be used to calculate the document-topic distribution representation R_C of the input short text. CRFTM is capable of discovering the document-topic distributions of the whole dataset in advance, and only needs to be trained once, without increasing the training time of the model. Since the inference of the topic model and the forward propagation of contextualized representations are decoupled, this is easy to achieve in the TDDM model. By combining the topic-based representation R_C with the contextualized representation R_X of the short text, we can obtain the final text feature representation R_{text}:

$$R_{text} = R_C \oplus R_X , \tag{4}$$

where $\oplus$ stands for feature concatenation operator.

3.3 User Level Features

In social media platforms, there are various features to represent the overall behavior of users. Unlike post level features extracted from a single post, user level features are usually obtained from all posts. The proposed method extracts the following user-level features for each user:

Sentiments. Emojis have become a common tool for communication in social media, because users can express their emotional state with simple symbols. Distinguishing between positive and negative emotions is very helpful for understanding the semantics of short texts [16]. Therefore, the proposed model first counts the number of negative, positive and neutral emojis appearing in each short text. We then sum up the number of emojis in all posts of each user to obtain the number of positive, neutral, and negative emojis at the user level. Furthermore, Valence Aware Dictionary and sEntiment Reasoner (VADER)[3] is utilized to extract users' emotional features.

User Information. We extract several features from user data, which can reflect the unique information of each user. Features such as gender, age, the number of friends and the number of followers can be obtained from the user account. In addition, we also collected the number of tweets and reposts of each user. For each user, we treat gender as a binary vector, and other features as one-hot labels among five intervals. The time distribution of tweets posted every day is counted and represented by a 24-dimensional vector.

Linguistic Inquiry and Word Count (LIWC). LIWC is an analysis tool that counts the psychologically related words in the text, which can be used to analyze the user's mental state from tweets [17]. Many studies have shown that when LIWC is applied to Twitter textual content, there is a correlation with depressive symptoms. Following [18], we selected eight LIWC features and analyzed all tweets of each user. Furthermore, we collected a dictionary containing antidepressants from Wikipedia, and counted the number of times these words appeared in user tweets.

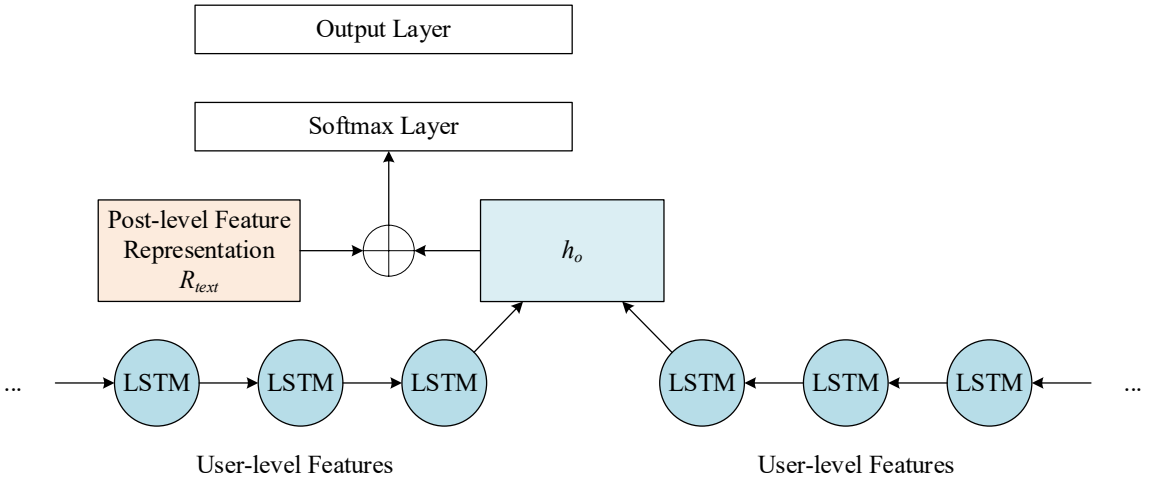

Fig. 3. The overall structure of TDDM

3 https://github.com/cjhutto/vaderSentiment/

TDDM employs a joint learning mechanism to fuse multiple features to detect depressed users, as shown in Fig. 3. Recurrent Neural Network (RNN) is a network model that sequentially links hidden layer nodes and can dynamically extract sequence features. Although vanilla RNN can convey the semantic relationship between words, it is difficult to capture long-distance information. Using the mechanism of forget gate f, output gate o and input gate i, LSTM is able to effectively capture long-distance connections between user features. However, one-way LSTM only captures the forward connection of features, and we use BiLSTM to obtain the forward and backward relationships between user-level features.

The hidden layer of BiLSTM has a backward output and a forward output, which are defined as:

$$\begin{aligned} \overleftarrow{h_t} &= LSTM_{backward}(h_t, \overleftarrow{h_{t-1}}) \\ \overrightarrow{h_t} &= LSTM_{forward}(h_t, \overrightarrow{h_{t-1}}) \\ h_o &= \overleftarrow{h_t} \oplus \overrightarrow{h_t} \end{aligned}, \tag{4}$$

where h_t is the hidden layer state of the LSTM at the current moment, $\overleftarrow{h_t}$ and $\overrightarrow{h_t}$ indicate the hidden layer output in the backward and forward directions, respectively. Let user-lever features be represented by x, and the calculation process of BiLSTM is as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t]) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t]) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t]) \\ \hat{m}_t &= tanh(W_m \cdot [h_{t-1}, x_t]) \\ m_t &= f_t \cdot m_{t-1} + i_t \cdot \hat{m}_t \\ h_t &= o_t \cdot tanh(m_t) \end{aligned}, \tag{5}$$

where σ denotes a sigmoid activation function, W represents the parameters of TDDM. After encoding the post-level and user-lever features, TDDM concatenates them and feeds them to the final softmax layer for classification.

4 Experiments

In this section, we introduce experiments to evaluate the performance of the TDDM model. The performance in terms of precision, recall and F1-measure is reported on a real-world Twitter dataset.

4.1 Dataset

We first reuse the dataset provided by [3], which contains large-scale real Twitter data with reliable labels. When they retrieved a signal tweet to infer that the user is suffering from depression, they collected the user's profile data. According to doctors' recommendations, the diagnostic criteria for depression are related to the duration of symptoms, so they also collected all tweets posted by users within a month. Furthermore, we utilize Tweety API to collect user information and tweets of depressed users during the COVID-19 pandemic. Following [3], once phrases such as "I have depression" or "I got depression" appear in a tweet, we think the user is suffering from depression.

In the experiment, the Natural Language Processing Toolkit (NLTK)[4] is used for data cleaning, which is a Python library for data preprocessing tasks such as removing stop words and part-of-speech tagging. Since words such as "the" or "a" are not helpful for TDDM to understand the semantics of tweets, with the help of NLTK, all stop words in the corpus are removed. A large number of stop words increase the dimensionality of textual features, resulting in high temporal and spatial complexity of the model, and reducing the performance of depressed

4 http://www.nltk.org/

user recognition. Furthermore, we remove the non-alphabetic characters and unnecessary spaces to further filter the noise in the dataset.

The above data preprocessing process can improve the quality of the corpus, and enhance the performance of TDDM. In addition, the efficiency of the model has been improved due to the reduction of the dimensionality of the corpus. Finally, we obtained 2153 depressed users and 2200 ordinary users.

4.2 Baseline Methods

In the experiment, TDDM is compared with the following models:

CNN. CNN is a neural network model that uses convolutional filters to extract local features from data [30]. Initially, the CNN model was mainly used in computer vision. However, CNN was later proved to be effective for NLP, and achieved excellent results in tasks such as text classification and automatic summarization.

BiLSTM. BiLSTM is a sequence-to-sequence neural network model, which is composed of two LSTMs: one for backward processing and the other for forward processing. The advantage of BiLSTM is that it increases the amount of information available to the algorithm, and improves the model's ability to capture contextual semantics.

BERT. Structurally, BERT is a transformer-based encoder. Transformer is an encoder-decoder neural network that exploits a self-attention mechanism in the encoder [13]. The BERT model is first pre-trained on a large-scale dataset, and then an output layer is added to fine-tune the pre-trained model.

ALBERT. ALBERT is a lightweight architecture based on BERT, and its parameters are far fewer than those of the BERT model. ALBERT integrates two parameter reduction methods: cross-layer parameter sharing and a factorized embedding parameterization mechanism, alleviating the problem of the huge amount of parameters of the pre-training model.

XLNet. XLNet is an autoregressive pre-trained neural network model, which takes full advantage of autoencoding and autoregressive modeling, and tries to solve their limitations [10]. Furthermore, by incorporating the relative coding mechanism and segment recursive scheme into the pre-training process, XLNet has achieved better performance on tasks of long text sequences.

For CRFTM, Dirichlet priors α and β are set to 0.5 and 0.01 respectively. The number of Gibbs sampling iterations is set to 1000, and the other parameters of the topic model are consistent with the original paper. For the two models of BiLSTM and CNN, we use the freely-available word2vec tool[5], which provides a 300-dimensional word vector for each word. We set the dropout value to 0.2, which helps reduce the interdependence between neurons.

We randomly select 80% of the data from the dataset for training, and the rest form a test set. All models employ the cross-entropy loss function in the training process. The adam optimizer is used for training, and its learning rate is 2e-5. For BERT and XLNet, the BERT-Base model[6] and the XLNet-Base[7] model are used to detect depressed users, where the size of the hidden layer and the number of self-attention heads are 12 and 768, respectively. For ALBERT, we utilize the second version of the basic model for experiments, where the hidden dimension and attention head are set to 768 and 12 respectively.

4.3 Evaluation Metrics

F1-measure, precision and recall are used to evaluate the performance of TDDM and baseline models. Three important terms related to evaluation metrics are as follows:

- TP: The number of real depressed users identified as depressed.
- FP: The number of real depressed users identified as non-depressed.
- FN: The number of ordinary users identified as depressed.

F1-measure, precision and recall can be calculated as follows:

5 http://code.google.com/p/word2vec/
6 https://github.com/google-research/bert/
7 https://github.com/zihangdai/xlnet/

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1 &= \frac{2\times Precision\times Recall}{Precision+Recall} \end{aligned} . \tag{6}$$

4.4 Experimental Results

Jamil et al. pointed out that just one tweet is not enough to determine whether a user is suffering from depression [19]. Therefore, we concatenate the tweets of the same account to form pseudo-documents containing 200 words, and label these pseudo-documents according to the tags of the accounts. In Fig. 4, we report the experimental results of the TDDM and baseline models at the pseudo-document level. An immediate observation is that TDDM achieves the best results for pseudo-document-level depression detection compared to baseline methods.

Another observation is that the experimental results of depressed user detection based on pre-training models outperform traditional neural network models such as CNN and BiLSTM. The reason is that these models employ the transformer architecture to pre-train on a large external corpus, and capture common language representations to improve the performance on downstream tasks. The performance of CNN is slightly better than BiLSTM, which may be because CNN is better at extracting local and location invariant textual features, and is more suitable for the dataset used in this paper. Although the parameters of the ALBERT model are significantly less than those of the BERT model, the performance of ALBERT is still very competitive. The XLNet model achieves the second best experimental result. The reason may be that it learns contextual information through various permutations of the input sequence.

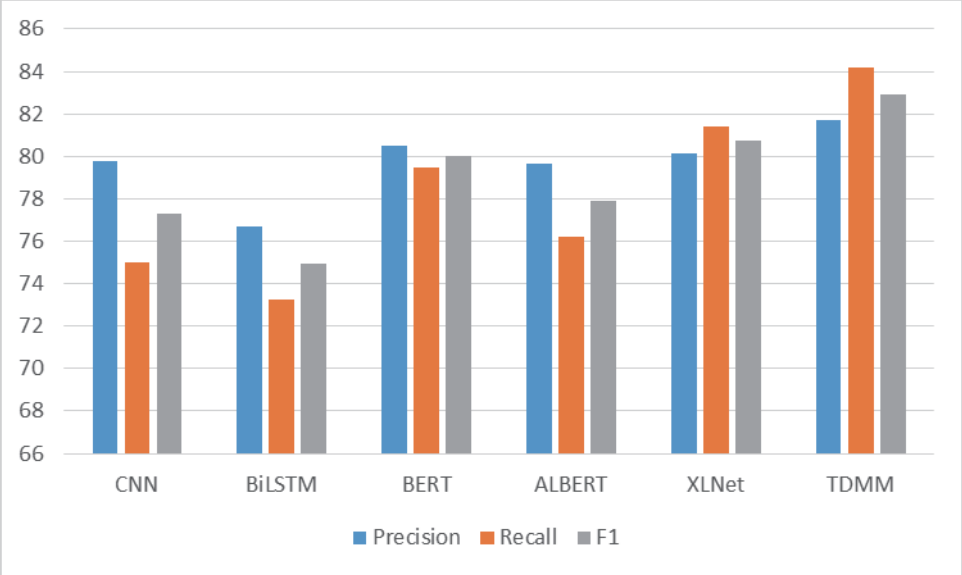

Fig. 4. Pseudo-document level depressed user detection

The TDDM model integrates topic knowledge into XLNet, which makes its experimental results outperform other baseline models. This topic information extends the textual characteristics of tweets, and provides a perspective for the model to understand text collections. Furthermore, TDDM utilizes a series of features to learn user-level depression-related behaviors, thus forming a robust model.

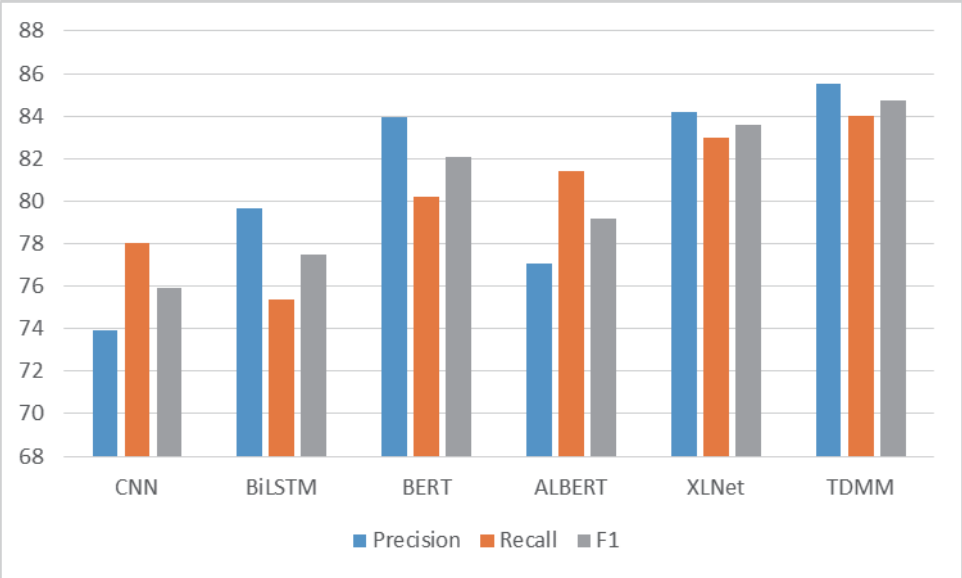

Fig. 5. User-level depressed user detection

In Fig. 5, we report the experimental results of different models on user-level depression detection. As shown in the figure, compared with the experimental results of pseudo-document level, the performance of all models has been improved at the user level. The experimental results show that more text content has a positive effect on the model to determine whether the user is depressed.

Fig. 6 shows the classification results of the ablation study, validating the contribution of the TDDM model. "Topic Feature Only" denotes that only topic features are used to predict depressed users in social media, and XLNet is used to encode posts. "User Level Feature Only" represents that topic features are not used in the classification process, and XLNet extracts textual features.

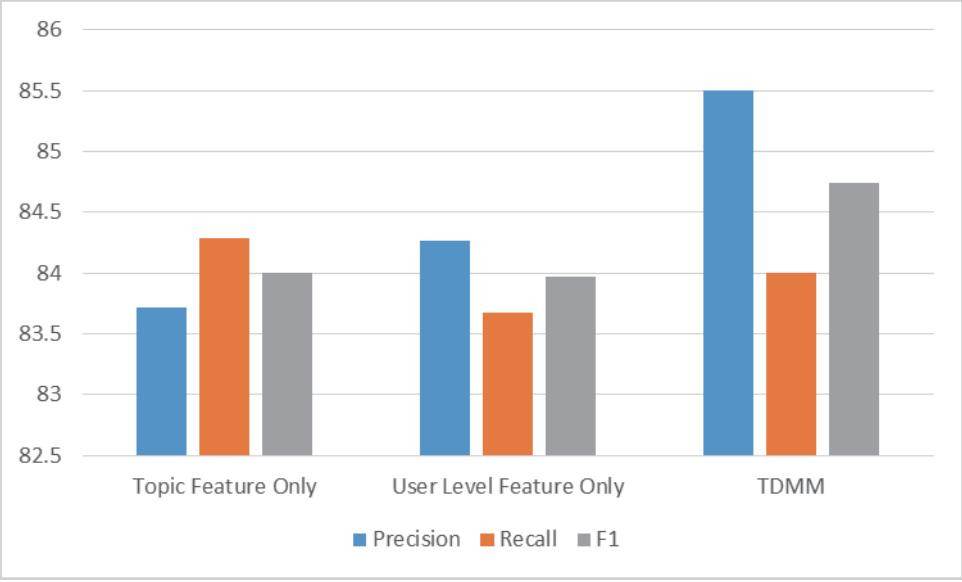

Fig. 6. Ablation study

It can be seen from the results that on the basis of the XLNet model, each module of TDMM has achieved a performance improvement of 4-5%. Another observation is that topic features and user level features can promote each other to improve the performance of depression detection in social media.

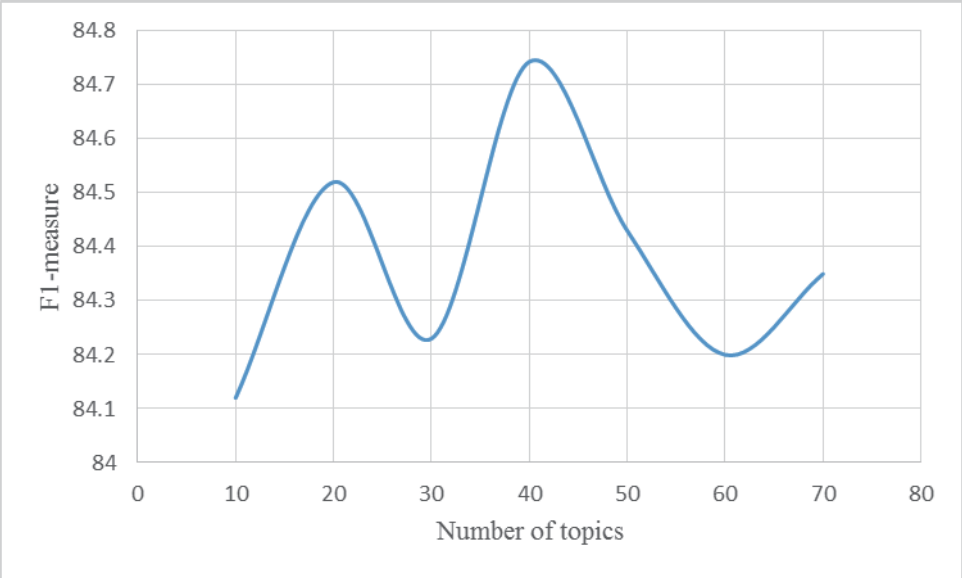

Fig. 7 Results of TDDM with different number of topics

Following [33], we experimentally explore the effect of the number of topics in TDDM. The F1-measure of TDDM for different number of topics is shown in Fig. 7. When the number of topics is 40, our model achieves the best experimental result. Because too many topics lead to overfitting, the F1-measure of TDDM decreases as the number of topics continues to increase.

5 Conclusion

This paper proposes a new model to predict depressed users in social media platforms by generating features from posts and user behaviors, which is called the Topic-enriched Depression Detection Model (TDDM). TDDM divides the prediction process of depressed users into three steps. Our model first utilizes a Conditional Random Field Regularized Topic Model (CRFTM) to extract the topic information of posts, and obtain their topic distributions. Secondly, we learn the textual representation of posts from XLNet and topic distributions. Finally, we incorporate user level features into TDDM to further improve detection performance. Experiments on a real-world Twitter dataset verify the effectiveness of TDDM in depression detection. In the future, we will mine the characteristics of pictures posted by users to improve our model's performance in predicting depressed users. In addition, we will also study how to apply TDDM to other NLP tasks, such as fake news detection or opinion mining.

Acknowledgments

We would like to thank the anonymous reviews for their valuable comments. This work was supported in part by the Scientific Research Program of Hubei Provincial Department of Education (No. B2021063).

References

[1] R. Kessler, M. Angermeyer, J. Anthony, R. Graaf, K. Demyttenaere, I. Gasquet, G. Girolanmo, S. Gluzman, O. Gureje, J.M. Haro, N. Kawakami, A. Karam, D. Levinson, M.E. Mora, M.O. Browne, J. Posada-Villa, D.J. Stein, C.H. Tsang, S. Aguilar-Gaxiola, J. Alonso, S. Lee, S. Heeringa, B. Pennell, P. Berglund, M. Gruber, M. Petukhova, S. Chatterji, T.B. Ustün, Lifetime prevalence and age-of-onset distributions of mental disorders in the who world mental health (WMH) surveys, World Psychiatry 6(2007) 168-176.

[2] P.N. Chukwueke, The relationship between mental illness and mass shooting, [Ph.D. dissertation] Minneapolis: Walden University, 2020.

[3] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T. Chua, W. Zhu, Depression detection via harvesting social media: A multimodal dictionary learning solution, in: Proc. International Joint Conference on Artificial Intelligence, 2017.

[4] M. Park, C. Cha, M. Cha, Depressive moods of users portrayed in twitter, in: Proc. ACM SIGKDD Workshop on Healthcare Informatics, 2012.

[5] T. Gui, L. Zhu, Q. Zhang, M. Peng, X. Zhou, K. Ding, Z. Chen, Cooperative multimodal approach to depression detection in twitter, in: Proc. Conference on Artificial Intelligence, 2019.
[6] C. Chiu, H.Y. Lane, J. Koh, A.L.P. Chen, Multimodal depression detection on instagram considering time interval of posts, International Journal of Intelligent systems 56(1)(2021) 25-47.
[7] P. Resnik, W. Armstrong, L.M.B. Claudino, T. Nguyen, V. Nguyen, J.L. Boyd-Graber, Beyond LDA: exploring supervised topic modeling for depression-related language in twitter, in: Proc. Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015.
[8] R. Bagby, D. Schuller, A. Levitt, R. Joffe, K. Harkness, Major depression and the five-factor model of personality, Journal of affective disorders 38(1996) 89-95.
[9] W. Gao, M. Peng, H. Wang, Y. Zhang, Q. Xie, G. Tian, Incorporating word embeddings into topic modeling of short text, Knowledge and Information Systems 61(2019) 1123-1145.
[10]Z. Yang, Z. Dai, Y. Yang, J.G. Carbonell, R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Proc. Annual Conference on Neural Information Processing Systems, 2019.
[11]J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: Proc. Conference on Empirical Methods in Natural Language Processing, 2014.
[12]T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proc. Annual Conference on Neural Information Processing Systems, 2013.
[13]J. Devlin, M. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.
[14]W. Gao, L. Li, X. Zhu, Y. Wang, Detecting disaster-related tweets via multimodal adversarial neural network, IEEE Multimedia 27(4)(2020) 28-37.
[15]Z. Dai, Z. Yang, Y. Yang, J.G. Carbonell, Q.V. Le, R. Salakhutdinov, Transformer-xl: Attentive language models beyond a fixed-length context, in: Proc. Conference of the Association for Computational Linguistics, 2019.
[16]Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, C. Sun, Facebook sentiment: Reactions and emojis, in: Proc. International Workshop on Natural Language Processing for Social Media, 2017.
[17]Y.R. Tausczik, J.W. Pennebaker, The psychological meaning of words: Liwc and computerized text analysis methods, Journal of Language and Social Psychology 29(1)(2021) 24-54.
[18]G. Coppersmith, M. Dredze, C. Harman, Quantifying mental health signals in twitter, in: Proc. Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2004.
[19]Z. Jamil, D. Inkpen, P. Buddhitha, K. White, Monitoring tweets for depression to detect at-risk users, in: Proc. Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality, 2017.
[20]M.D. Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, in: Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), 2013.
[21]Q. Hu, A. Li, F. Heng, J. Li, T. Zhu, Predicting depression of social media user on different observation windows, in: Proc. International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015.
[22]E.A. Ríssola, M. Aliannejadi, F. Crestani, Beyond modelling: Understanding mental disorders in online social media, in: Proc. of European Conference on IR Research (ECIR), 2020.
[23]T. Shen, J. Jia, G. Shen, F. Feng, X. He, H. Luan, J. Tang, T. Tiropanis, T. Chua, W. Hall, Cross-domain depression detection via harvesting social media, in: Proc. International Joint Conference on Artificial Intelligence (IJCAI), 2018.
[24]M. Wu, C. Shen, E.T. Wang, A.L.P. Chen, A deep architecture for depression detection using posting, behavior, and living environment data, Journal of Intelligent Information systems 54(2)(2020) 225-244.
[25]J. Hussain, F.A. Satti, M. Afzal, W.A. Khan, H.S.M. Bilal, M.Z. Ansaar, H.F. Ahmad, T. Hur, J. Bang, J.-I. Kim, G.H. Park, H. Seung, S. Lee, Exploring the dominant features of social media for depression detection, Journal of Information science 46(6)(2020) 739-759.
[26]L. Genevieve, D. Huang, W. Lin, Context-aware deep learning for multi-model depression detection, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2019.
[27]A.H. Orabi, P. Buddhitha, M.H. Orabi, D. Inkpen, Deep learning for depression detection of Twitter users, in: Proc. Workshop on Computational Linguistics and Clinical Psychology (CLPsych), 2018.
[28]D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, Journal of Machine Learning Research 3(2003) 993-1022.
[29]D.M. Blei, J.D. Lafferty, Dynamic topic models, in: Proc. International Conference on Machine Learning (ICML), 2006.
[30]Y. Kim, Convolutional neural networks for sentence classification, in: Proc. Empirical Methods in Natural Language Processing (EMNLP), 2014.
[31]W. Gao, M. Peng, H. Wang, Y. Zhang, W. Han, G. Hu, Q. Xie, Generation of topic evolution graphs from short text streams, Neurocomputing 383(2020) 282-294.
[32]B. Du, G. Liu, Topic analysis in LDA based on keywords selection, Journal of Computers 32(3)(2021) 1-12.
[33]W. Gao, Y. Fang, L. Li, X. Tao, Event detection in social media via graph neural network, in: Proc. Web Information Systems Engineering (WISE), 2021.