Suicide Risk Assessment on Social Media

Yanfei Huang
Faculty of Electronic Information and Electrical Engineering, Dalian University of technology, Dalian, China

Keywords: Suicide, Media, BiLSTM, Attention.

Abstract: Suicide is a global problem, and the number of people suffering from suicidal ideation is increasing globally. Therefore, suicide risk assessment is critical. With the development of the Internet in recent years, social media has become an essential source of information for studying psychological disorders such as depression and suicide.

To this end, this paper designed a two-layer BiLSTM attention network model, using users' posts on social media as input to assess users' suicidal ideation levels. In order to improve the performance of the model, this paper sorted the posts according to the time stamp when preprocessing the dataset and also used the pre-training language model BERT, which can obtain a more reasonable word vector representation than the word embedding model.

This paper assesses this model on the dataset provided by CLPsych2019. The dataset was taken from Reddit and divided users into four categories: no, low, moderate, and severe. The final experimental results show that the Accuracy of the model proposed in this paper can reach 62%, and the Macro_F1 value reaches 0.438. So, the model is a suitable assessment method.

1 INTRODUCTION

1.1 Research Aim and Significance

Suicide has become a global issue. According to the World Health Organization (WHO), 75% of suicides occurred in low-and middle-income countries in 2012. In the same year, suicide was the second major cause of death among people aged 15-19. Every year, more than 800,000 people die from suicide, and more have suicidal thoughts. Although China had had the lowest suicide rate by the end of 2011, 9.7people died of suicide per 1,000,000 people in 2016. Therefore, suicide prediction and prevention are of the essence.

1.2 Research Status at Home and aboard

Suicide risk assessment is a kind of text classification task. With the development of machine learning, methods for suicide risk assessment are also increasing with higher prediction rates. As a sub-class of machine learning, deep learning's complexity can feature raw data extraction and potentially find better solutions. However, the discussion of deep learning and suicide risk assessment is limited. Therefore, it is crucial to address deep learning algorithms and their applications in suicide risk assessment.

Deep Learning

The main deep learning architectures for text classification include Rule-embedded Neural Networks (ReNN), Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN) (Khalil Alsmadi, Omar, Noah et al., 2009). Deep learning models are believed to have better performance in text classification. Ji (Ji, Yu, Fung et al., 2018) and his research fellows have compared the other five machine learning models with LSTM, and verified the feasibility and practicability of these models. Their study has laid a significant foundation for suicide assessment on Reddit and Twitter. Kalchbrenner et al. (Kalchbrenner, Grefenstette, Blunsom, 2014) have proved that the CNN-based method has advantages on N-gram features. Cao (Cao, Zhang, Feng, 2020) has utilised the personal knowledge map (PKM) for suicide prediction with an Accuracy of 93.74%. Their study applied a two-layer attention mechanism (Attention mechanism and neighbour Attention) to find out critical indicators of suicide ideation.

Huang, Y.
Suicide Risk Assessment on Social Media.
DOI: 10.5220/0011752000003607
In *Proceedings of the 1st International Conference on Public Management, Digital Economy and Internet Technology (ICPDI 2022)*, pages 560-564
ISBN: 978-989-758-620-0

1.3 Summary

In conclusion, deep learning performs better than traditional machine learning methods in suicide assessment. Therefore, this paper will mainly focus on deep learning models, such as CNN, RNN, and LSTM. Derived from RNN, LSTM can capture long-term dependencies and alleviate the common problems of gradient disappearance and gradient explosion in RNN. Moreover, LSTM will be more suitable for suicide prediction as many long posts on social platforms might appear. As a result, this paper tends to reproduce the LSTM model and study BERT (Bidirectional Encoder Representations from Transformers) word embeddings and Attention to improve the Accuracy of suicidal ideation detection on social platforms and analyse the results.

2 DATASETS AND EVALUATION INDEX

2.1 Dataset

The dataset applied in this study is from the shared task for the 2019 Workshop on Computational Linguistics and Clinical Psychology, known as CLPsych 2019, whose goal is to assess the risk of suicide according to users' posts on Reddit. The database collected posts include the poster's ID, user ID, timestamp, subreddit, post title, and post body. Each ID corresponds to a suicide level, stored in a separate CSV file with two columns—post ID and Label, to identify users at no, low, moderate, or severe risk. This paper will implement a four-classification task based on CLPsych 2019, and the dataset is shown in the table below.

2.2 Evaluation Metrics

In this four-classification task of suicide risk assessment, this paper applies two evaluation metrics: Accuracy (Acc) and Macro_F1. On this basis, two classifications are added: the presence or absence of suicidal ideation and the severity of suicidal ideation. By analysing Accuracy (Acc), Precision (P), and F1 score, this paper will provide a comprehensive view of strengths and weaknesses among models to analyse the reasons for the changes in their effectiveness.

Thanks to scikit-learn, the Python machine learning library, the author could obtain scores, including Recall (R), Accuracy (Acc), F1, and the Macro_F1 on multi-task classification by inputting y_pred and y_true and calling the corresponding functions.

3 THE PREDICTION METHOD BASED ON TWO-LAYER BILSTM MODEL

This Chapter will discuss in detail the suicide risk assessment method based on the two-layer BiLSTM model, including the research goal and objective, principle, method, the setting of BiLSTM parameters, and cause analysis.

First, the objective of this experiment is shown in Figure 3.1 below. Posts on social media are input into the Neural Network and obtain a label reflecting the level of suicidal ideation with four classifications (no, low, moderate, or severe). The main research goal is to obtain the predictions in the MyNet part and evaluate the model using the evaluation metrics mentioned above.

Table 1: Format of the Dataset [Owner-draw].

Post_ID	User_ID	timestamp	subreddit	Post_title	Post_body
wfimt	22002	1342075703	League oflegends	Scared of next ban. What can I do?	Hi, Guys. I already got a perm ban…
2ddokl	22002	1407882753	No Fap	Am depressed,At no fap,and a bitdrunk.	I want to say everything... thanks that you are here... Pls chat with me! i will chare my exp with no fap!
15c0je	22014	1356285627	SuicideWatch	I don't think I can go on anymore.	I am a failure. I am experiencing the worst body dysphoria of my life.

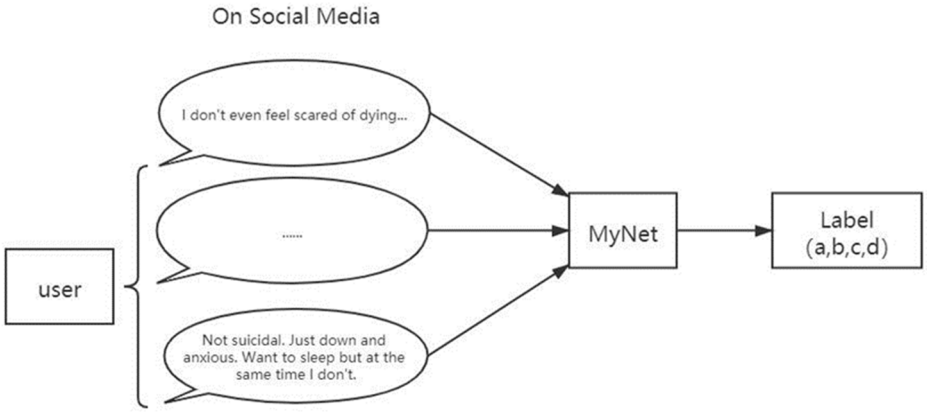

Figure 1: Experimental Objectives [Owner-draw].

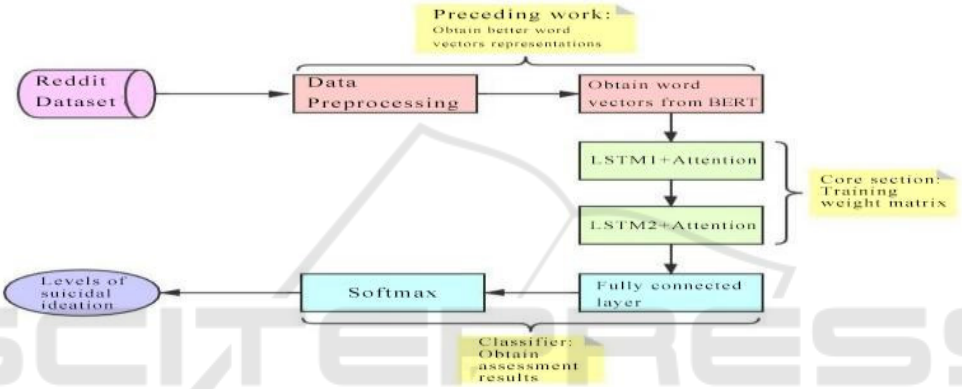

Figure 2: The General Workflow [Owner-draw].

The design of MyNet is based on the BiLSTM model, whose workflow is shown below (Figure 3.2). After getting access to the dataset, we first need to preprocess the data, remove noisy data and other meaningless information, and obtain the word vector through BERT for better representation. Then, the word vector is passed through the LSTM1 and weighted by the Attention to find connections of words in the same post. Next, it is input into LSTM2 and weighted by the Attention. Last, the result is put into the fully connected layer and normalised through the Softmax layer to get the probability of each label.

3.1 BERT Word Embeddings

After preprocessing the data, we can get the content of the posts sorted by time. The posts are still in characters. However, BiLSTM only recognises digital data. On the one hand, one-hot encoding cannot represent the connections among words, and word2vec does not handle words with multiple meanings. On the other hand, BERT is more suitable to obtain word vectors as it can dynamically represent the word in accordance with the context.

3.2 The Two-Layer BiLSTM with Attention

The word embeddings give a 768-dimension vector for each word in the posts on Reddit. Then the results are fed into the neural network for matrix training to find the connection between the input tensor and suicidal ideation level.

Long Short-Term Memory

LSTM (Long short-term memory) is developed from RNN, which will suffer from the problem of vanishing and exploding gradients in long data sequences. LSTM, on the other hand, has overcome these shortcomings. To assess suicide risks, we need to process many long posts, so the LSTM model is more suitable than other models.

- **input** of shape (*seq_len, batch, input_size*): tensor containing the features of the input sequence. The input can also be a packed variable length sequence. See `torch.nn.utils.rnn.pack_padded_sequence()` or `torch.nn.utils.rnn.pack_sequence()` for details.

Figure 3: LSTM Input Format [Owner-draw].

Attention in Machine Learning

Attention can be attached to Encoder-Decoder and other models as a mechanism to simulate cognitive Attention. Thus, the author employs the Attention mechanism in the model. As long-term dependencies might cause gradient explosion in LSTM, a self-attention mechanism can remove meaningless information to avoid exploding and vanishing Gradients. The formula for Attention is as follows:

$$Attention\ (Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})\ V \qquad (1)$$

There are two advantages of Attention. First, it concerns all the information from the word in the posts. Second, it can perform parallel computing, improving the operation's efficiency.

BiLSTM-Attention and Its Implementation

Bidirectional LSTM (BiLSTM) is a recurrent neural network mainly used for natural language processing. Unlike standard LSTM, the input flows in both directions and can utilise information from both sides. Table shows that LSTM is single-directional. When inputting timestep t, we can only consider the information before it, whereas BiLSTM processes the information from front to back and the information from back to front. Therefore, each timestep in BiLSTM will produce two hidden states. If we input the hidden dimension as 100, the output hidden size will be 100*2=200. Here is the defined value of LSTM in PyTorch.

Thus, the author sets up a two-layer BiLSTM model. The first layer is designed to gain the vector representation of posts, and the second layer is set up to get the representation of users. Assuming that there are seven posts of a user with 100 words for each, the hidden dimension assumed is 100 in this LSTM, in other words, ht = 100, and here is the architecture of the model:

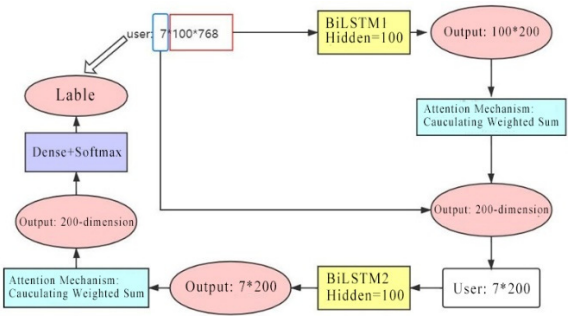

Figure 4: The Architecture of the Two-layer BiLSTM-Attention Model [Owner-draw].

3.3 Results and Analysis

The value of Loss Function and Accuracy (Acc) is diagrammed as follows:

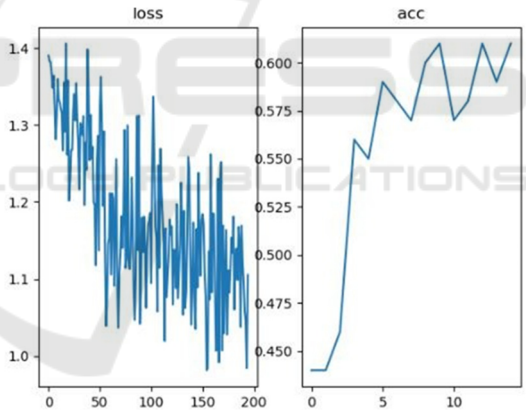

Figure 5: Experimental results of double-layer BiLSTM attention network[Owner-draw].

The abscissa of the left figure is the number of times, and the ordinate is the loss value; the abscissa of the right figure is the training round times (epoch), in total 15, and the ordinate is the prediction accuracy. The figure shows that the Loss value is declining, and the Acc value is rising with the training rounds increase, a peak at 62%. See the following table for other evaluation indicators:

Table 2: Evaluation Metrics of the Double-layer BiLSTM-Attention Model [Owner-draw].

Classifications	Recall	Precision	Accuracy	F1-measure
Risk（0，1，2，3）	/	/	62%	0.438
Suicidal ideation Presence（0，1）	0.948	0.924	90%	0.936
Severe Suicidal Ideation（0，1）	0.956	0.835	84%	0.892

The above table shows that the model can classify the presence of suicidal ideation with high Recall and Accuracy. In contrast, the Accuracy has declined by nearly 10% in determining the severity of suicidal ideation. Therefore, the model in this paper cannot distinguish the slight differences between the low and moderate levels. Moreover, it usually classifies users with low suicidal ideation as moderate.

While this model performed well in suicidal ideation classification, the Accuracy and F1_macro value are relatively lower in the four-classification task due to the model's design and dataset applied. In CLPsych2019, fewer users without suicidal ideation and more in the remaining three categories. Thus, it is easier to determine suicidal ideation as the accuracy rate is higher under this classification criterion.

4 CONCLUSION

The task of this paper is to collect posts on social media and classify them into four categories: no, low, moderate, and severe, according to the level of suicidal ideation.

To complete this task, the author set up a two-layer BiLSTM-Attention Model. The model first cleans the data, removing noisy and meaningless data. Then it sorts each user's posts by timestamp and obtains the word vectors through BERT, the pre-trained language model, which provides a better representation than the word embedding model. Next, the results are fed into the model designed in this paper, utilising BiLSTM to capture long-distance dependency (LDD) and the bidirectional information as well as the Attention mechanism, which allows the model to focus on the core information of the post. The author obtains the reasonable vector representations of post through the first layer of BiLSTM, the representations of the users in the second layer, and the assessment results from the classifiers. The Accuracy of the model reaches 62%, and the macro-f1 value is 0.438. Specifically, the Attention mechanism improves the Macro_ Fi value by 16%. Compared with other models feeding the results directly to the classifiers after BERT, this model improves the Macro_F1 value by 29%. As a result, the model in this paper performs better on the processing dataset CLPsych2019.

What is more, to better assess the model's performance, the author introduces two classifications: the presence of suicidal ideation and the severity of suicidal ideation. Results show that the distinguishability of low and moderate suicidal risk is relatively low, which needs to be improved in the future.

REFERENCES

Maron M E. Automatic indexing: an experimental inquiry[J]. Journal of the ACM (JACM), 1961, 8(3): 404-417.

Rabiner L, Juang B. An introduction to hidden Markov models[J]. IEEE ASSP Magazine, 1986, 3(1): 4-16.

Khalil Alsmadi M, Omar K B, Noah S A, et al. Performance comparison of multi-layer perceptron (Back Propagation, Delta Rule and Perceptron) algorithms in neural networks[C]//2009 IEEE International Advance Computing Conference. IEEE, 2009: 296-299.

Ji S, Yu C P, Fung S, et al. Supervised learning for suicidal ideation detection in online user content[J]. Complexity, 2018.

Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.

Cao L, Zhang H, Feng L. Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media [J]. IEEE Transactions on Multimedia, 2020.