Suicidal risk identification in social media

Article *in* Procedia Computer Science · January 2021

DOI: 10.1016/j.procs.2021.05.106

Ashok Kumar Jayaraman
Anna University, Chennai

Tina Esther Trueman
Anna University, Chennai

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 189 (2021) 368–373

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on AI in Computational Linguistics

Suicidal risk identification in social media

Ashok Kumar J[a,*], Tina Esther Trueman[a], Abinesh A K[b]

[a]Department of Information Science and Technology, Anna University, Chennai-600025, India
[b]Department of Journalism, Madras Christian College, Chennai-600059, India

Abstract

Social media influences people to express their mental health issues such as depression and anxiety. Specifically, depression is one of the biggest risk factors for suicidal ideation and attempts. Therefore, we propose a multiplicative attention-based bidirectional gated recurrent unit to identify the suicidal risk factors of social media users. The proposed model captures the local context in input sequences. Our experimental results indicate that the proposed model outperforms the state-of-the-art models in the multiclass classification task.

Peer-review under responsibility of the scientific committee of the 5th International Conference on AI in Computational Linguistics.

Keywords: Behavioral monitoring; suicidal ideation; deep learning; gated recurrent unit; attention mechanism.

1. Introduction

In modern society, suicide is rapidly increasing due to mental health problems such as anxiety and depression [1]. Anxiety is a normal feeling or an emotion where the brain reacts to the stress and alerts potential danger ahead. For instance, problems at work, making an important decision, and fear of an activity or a situation. Depression is associated with a feeling of sadness, loss of interest, or anger of an individual. In particular, the World Health Organization indicated that suicide is the second largest cause of death worldwide among teenagers [2, 3]. Nowadays, online social media influence individual users to express their feelings or emotions in the form of posts or comments [4, 5]. These personal feelings help us to identify suicidal risk factors, namely, ideation, indicator, behavior, attempt, and supportive [6]. First, the suicidal ideation category defines a suicidal thought of a user due to the loss of a strong relationship, loss of a job, mental illness, substance abuse, or chronic diseases. Second, the suicidal behavior involves actively planning to commit suicide, self-harm activity, using blunt force violence, or actions of death. Third, the attempt category is defined as a complete attempt, changed their mind, or writing a good-bye message. Fourth, suicidal indicator category involves at-risk language from acute symptoms, engagement in a supportive manner, history of divorce, chronic ill-

* Corresponding author: Ashok Kumar J
E-mail address: jashokkumar83@auist.net

1877-0509 © 2021 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 5th International Conference on AI in Computational Linguistics.
10.1016/j.procs.2021.05.106

ness, or sharing personal history. Finally, the supportive category involves active engagement without any history of risk in the past or present language.

Particularly, the participation of teenagers in online social media games involves self-harming activities [7]. Therefore, we need an automated system to identify suicidal risk factors in social media. The recent development of natural language processing (NLP), machine learning, and deep learning [8] helps us to develop such kind of system. Traditional machine learning models capture a bag of word features and deep learning models capture semantic context features in a fixed dimension. The fixed dimension techniques fail to remember longer input sequences. To resolve this issue, an attention mechanism has been proposed to deal with NLP applications [9]. The attention mechanism can be implemented to capture global context features and local context features [10]. Specifically, additive and multiplicative attention mechanisms represent context vectors in similar complexity. However, the multiplicative attention mechanism provides fast computation in practice with more space-efficient. Therefore, in this paper, we propose a multiplicative attention-based bidirectional gated recurrent unit to identify suicidal risk factors in social media. This paper contributes to the following.

- Identifies suicidal risk factors of online users.
- Employs a multiplicative attention-based bidirectional gated recurrent unit.
- Outperforms the state-of-the-art models for suicidal risk factor identification task.

The rest of this paper is structured as follows. Section 2 explains the related works in suicidal ideation. Section 3 presents the proposed multiplicative attention-based bidirectional gated recurrent unit model. In Section 4, results and discussion are presented. Finally, we conclude this work in section 5.

2. Related works

In this section, we present the recent works in suicidal ideation using user-generated content. In particular, Ji et al. [4] presented a supervised learning approach to identify suicidal ideation in user-generated contents. The authors indicated that the XGBoost algorithm achieves a better AUC score with the combination of statistical, topic, TF-IDF, POS, and LIWC features. Gaur et al. [6] studied a domain-specific learning framework to predict the suicidal risk factor of an individual. This study uses suicidal ontology and medical knowledge base to detect suicidal thoughts. Specifically, the authors developed a gold standard Reddit dataset of 500 users. Then, they performed suicide lexicon-based classification on two input forms: textual features (I1) and characteristics and textual features (I2). The authors indicated that the CNN (convolutional neural networks) model achieves better performance than RF (Random Forest) and SVM (Support Vector Machine) models. Moreover, Kumar et al. [13] explored suicidal identification using tweets. The authors indicated that an ensemble-based random forest algorithm achieves 99% accuracy for predicting the suicidal thought tweets. Sawhney et al. [14] investigated deep learning architectures for suicidal ideation detection using tweets. This study reveals that the convolutional long short-term memory network achieves 81.2% accuracy and 82.7% F1-score. In summary, most of the researchers focused on suicidal ideation as a binary classification problem using user-generated content. However, there is no publicly available gold standard dataset except the Reddit C-SSRS suicide dataset. Therefore, we use this dataset to study the multiplicative attention-based bidirectional gated recurrent unit for identifying suicidal risk factors in a multiclass environment.

3. Multiplicative attention-based gated recurrent Unit

In this section, we present the multiplicative attention mechanism with a bidirectional gated recurrent unit (BiGRU_Mattn) model to identify the suicidal risk of an individual as shown in Fig.1. This figure discusses each component as follows namely, dataset, word representation, BiGRU, multiplicative attention mechanism, and output layer.

3.1. Dataset

We use the Reddit C-SSRS gold standard dataset for the task of suicidal risk identification. This dataset is prepared by Gaur et al. [6] for creating a competitive baseline. In particular, the Reddit dataset contains 500 users' posts where

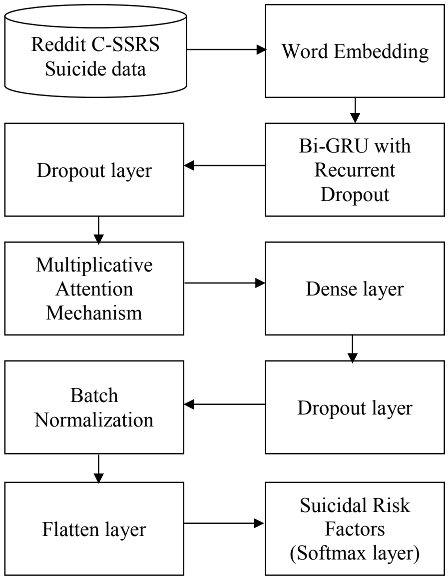

Fig. 1. The proposed Bidirectional GRU with multiplicative-attention network

each of these posts was annotated with one of the five different categories, namely, supportive, ideation, behavior, attempt, and indicator. Each of these categories contains 108, 171, 77, 45, and 99 posts respectively.

3.2. *Word Representation*

We preprocess the Reddit posts to remove special characters and punctuations. We then generate word vectors using the pre-trained word embeddings for words in the training data. Specifically, we use the GloVe (Global Vectors for word representation) for word embeddings. The word embeddings capture the semantic and syntactic information in a geometric space. In particular, the GloVe method captures the word co-occurrence statistics that how often a word occurs in a context window size [16]. Moreover, the GloVe word embedding improves model performance.

3.3. *Gated Recurrent Unit (GRU)*

Recurrent Neural Network (RNN) fails to learn long-term dependencies due to the vanishing gradient problem. To solve this issue, Cho et al. [18] introduced a special kind of gating mechanism is called Gated Recurrent Unit (GRU). It is similar to the Long Short Term Memory (LSTM) unit, where it makes each recurrent unit capture long-term dependencies. Particularly, the GRU has two gates, namely, the update gate and reset gate. These gates decide what information to be passed or removed without memory blocks for predicting the output. Moreover, the GRU network learns only from the previous context. To overcome this issue, Schuster et al. [19] introduced bidirectional RNNs to learn from the previous and future contexts. Specifically, we use a bidirectional GRU (BiGRU) network which comprises two layers of hidden nodes. The second layer reverses the input sequence and sends it to the network. Therefore, learning a sequence in both directions leads to predict a more accurate result.

3.4. *Multiplicative attention mechanism*

The GRU model cannot pay selective attention to the key part of the suicidal risk identification task. To address this issue, Bahdanau et al. [9] introduced an attention mechanism concept to capture the important part of a sentence. This concept has gained more popularity in the field of neural machine translations, computer vision, sentiment analysis, image caption generation, and image classification [11]. Moreover, the attention mechanism concepts are proposed in two broad categories, namely, global attention and local attention [10]. The global attention weights all the input

Table 1. Model performance

Models	Macro F1	Micro F1	Weighted F1
NB	**0.1951**	0.2360	0.2169
LR	0.1850	0.2060	0.2081
SVM	0.1640	0.2720	0.2288
LSTM_Attn	0.1261	0.2700	0.1797
BiLSTM_Attn	0.1410	0.2960	0.1968
GRU_Attn	0.1633	0.2920	0.2236
BiGRU_Attn	0.1661	0.2960	0.2203
LSTM_Mattn	0.1608	0.2980	0.2187
BiLSTM_Mattn	0.1534	0.2873	0.2106
GRU_Mattn	0.1601	0.2930	0.2197
BiGRU_Mattn	0.1914	**0.3000**	**0.2437**

words and the local attention weights only a subset of input words in the sentence. Particularly, the global attention is expensive to translate longer sequences. Specifically, additive and multiplicative attention concepts represent similar complexity. However, the multiplicative attention mechanism provides more space-efficient and fast computation. Therefore, we propose a multiplicative attention mechanism (also called dot-product attention) with attention width to control the local context for the task of suicidal risk identification. It uses the inner product to compute the attention score in a faster and efficient way. The main advantage of this approach is to avoid the expensive computation time. Let $H = [h_1, h_2, \cdots, h_t]$ be the hidden vectors produced by the BiGRU layer. Let $E = [e_1, e_2, \cdots, e_t]$ be the multiplicative attention energies. Then, we calculate the attention weight vector (a) and its weighted output (r) as follows in (1)-(4).

$$M = tanh(H) \tag{1}$$

$$E = \sigma(x_t^T W_a M + b_a) \tag{2}$$

$$a = softmax(E) \tag{3}$$

$$r = Ha^T \tag{4}$$

3.5. Output layer

The result of the multiplicative attention layer is passed to a fully connected layer for generating model results. These results passed to a dropout layer for regularizing the network. Furthermore, a batch normalization layer is applied to standardize the inputs in the network. Finally, we flatten this representation to the softmax output layer [20] for predicting the suicidal risk of an individual.

4. Results and discussion

We have evaluated the proposed model on the Reddit C-SSRS gold standard dataset. In this dataset, 500 instances are annotated with a set of predefined label categories such as supportive, ideation, behavior, attempt, and indicator. This dataset is randomly split into 10 fold for training and testing. Each fold contains 450 user posts for training and 50 user posts for testing. Then, we implement the multiplication attention-based bidirectional gated recurrent unit model in a Google Colaboratory notebook. Moreover, we used the GloVe word embeddings to generate word vectors for the train data with 100 dimensions. We also applied the input sequence length with 150 time-steps, one BiGRU layer with 64 units and 0.2 recurrent dropout rate, one attention layer with an attention width of 15, one dense layer with 32 units, one batch normalization layer, one flatten layer and an output layer with a softmax activation function. A dropout of 0.2 is used to approximate the training of neural networks. In particular, the proposed BiGRU model was employed on multiclass categories with a multiplicative attention mechanism. This model performs well with the Adam optimizer and the categorical cross-entropy loss function. Table 1 shows the result of the proposed model along with other models. The logistic regression model achieves a lower F1-micro score (20.60%) and the Naïve Bayes achieves the second-lowest F1-micro score (23.60%). Overall, the BiGRU model with a multiplicative attention mechanism shows a higher F1-micro score (30.00%) than other models. Particularly, the Reddit posts dataset is generated and studied using the lexicon-based method for identifying suicidal risk factors [6]. The authors used CNN with the ConceptNet-based embeddings and characteristics and Textual features as input data. In this paper, we used the attention-based BiGRU with GloVe word embedding on textual features for the suicidal risk factor identification task. More specifically, the attention mechanism has shown breakthroughs in NLP tasks. Therefore, it is chosen to compare with other models. The attention mechanism can access the whole input sequence length and can select specific elements of that sequence. Therefore, the proposed approach performs well on the suicidal dataset. The main limitations of this approach are that only 500 users' are considered and performed in a fixed sequence length with textual information. For complete suicidal risk identification, we can include users' images, videos, and emoticons as features. Furthermore, the proposed approach is trained and validated on English posts only. Therefore, this approach may not equally perform well on bi-lingual or multilingual posts.

5. Conclusion

In this paper, we proposed a multiplicative attention-based bidirectional gated recurrent unit model to identify the suicidal risk factors of an individual in social media. The proposed model achieves better performance than other models in the multiclass classification problem. In future work, suicidal ideation can be detected based on gender and location in a large dataset.

Acknowledgements

This work was supported by the University Grants Commission (UGC), Government of India under the National doctoral fellowship.

References

[1] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems.

[2] Coppersmith, G., Leary, R., Whyne, E., & Wood, T. (2015, August). Quantifying suicidal ideation via language usage on social media. In Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM (Vol. 110).

[3] World Health Organization et al. (2014). Preventing suicide: A global imperative. World Health Organization.

[4] Ji, S., Yu, C. P., Fung, S. F., Pan, S., & Long, G. (2018). Supervised learning for suicidal ideation detection in online user content. Complexity, 2018.

[5] Kumar, A., Abirami, S., & Trueman, T. E. (2018). Sentiment mining approaches for big data classification and clustering. In Modern technologies for big data classification and clustering (pp. 34-63). IGI Global.

[6] Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., ... & Pathak, J. (2019, May). Knowledge-aware assessment of severity of suicide risk for early intervention. In The World Wide Web Conference (pp. 514-525).

[7] Sinha, P. P., Mishra, R., Sawhney, R., Mahata, D., Shah, R. R., & Liu, H. (2019, November). # suicidal-A multipronged approach to identify and explore suicidal ideation in twitter. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 941-950).

[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[9] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[10] Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

[11] Kumar, J. A., Abirami, S., Ghosh, A., & Trueman, T. E. (2019, December). A C-LSTM with Attention Mechanism for Question Categorization. In Symposium on Machine Learning and Metaheuristics Algorithms, and Applications (pp. 234-244). Springer, Singapore.

[12] Chiroma, F., Liu, H., & Cocea, M. (2018, July). Text classification for suicide related tweets. In 2018 International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 2, pp. 587-592). IEEE.

[13] Rajesh Kumar, E., Rama Rao, K. V. S. N., Nayak, S. R., & Chandra, R. (2020). Suicidal ideation prediction in twitter data using machine learning techniques. Journal of Interdisciplinary Mathematics, 23(1), 117-125.

[14] Sawhney, R., Manchanda, P., Singh, R., & Aggarwal, S. (2018, July). A computational approach to feature extraction for identification of suicidal ideation in tweets. In Proceedings of ACL 2018, Student Research Workshop (pp. 91-98).
[15] Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C. C. J. (2019). Evaluating word embedding models: methods and experimental results. APSIPA Transactions on Signal and Information Processing, 8.
[16] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
[17] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.
[18] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.
[19] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 45(11), 2673-2681.
[20] Asadi, B., & Jiang, H. (2020). On Approximation Capabilities of ReLU Activation and Softmax Output Layer in Neural Networks. arXiv preprint arXiv:2002.04060.