Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/42/20230785

Discussion of a method for analysing adolescent depression based on BiLSTM and Attentional Mechanisms

Chang HaoXiang

Hebei Normal University College of Mathematical Sciences, Shijiazhuang, 050000, China

18134158182@163.com

Abstract. The incidence of depression in adolescents has been commonplace, but the number of people diagnosed with depression in hospitals is not large, so it is important to analyze people's textual content in normal times to determine whether they are in a depressive mood, and then intervene in a timely manner. The article firstly introduces the research results of previous scholars in sentiment analysis, and then briefly introduces the LSTM(Long Short-Term Memory) model, explains the operation formula of the model, and points out that BiLSTM(Bi-directional Long Short-Term Memory) has greater predictive ability than the LSTM model by taking the before and after contexts of the sentences into full consideration, and by extracting bi-directional semantic features. After that, the principle of the Attention Mechanism and the application of its derivative models, Self-Attention Mechanism, Multi-Attention Mechanism and Cross-Attention Mechanism, in different scenarios are analyzed, and finally, the advantages and disadvantages of BiLSTM and the attention mechanism are compared in the prediction of textual sentiment, and it is proposed that the BiLSTM-Attention model can play a stronger performance in the analysis of textual sentiment.

Keywords: depression, content analysis, BiLSTM, Attention Mechanism.

1. Introduction

Depression is a common mental illness that the World Health Organization ranks as one of the most common disorders worldwide. Depression has a wide and long-lasting negative impact on an individual's mood, thinking and behavior. Adolescents are the future of the world and it is even more important to analyze depression in adolescents. However, since the boundaries of diagnosing depression are blurred and an adolescent who is suffering from depression is hard to recognize in time, the analysis of texts in adolescents' daily communication becomes especially important. If people are able to know whether adolescents are potentially depressed in their texts, they will be able to detect and intervene in a timely manner.

In natural language processing, Sentiment Analysis plays an essential role and it can be divided into article-level analysis, sentence-level analysis, and word-level and phrase-level analysis according to the different levels of research [1]. Common approaches consist of evaluating the polarity of sentences using dictionaries, along with applying rules of sentence grammar., or labeling the polarity of sentences using manual annotation methods, and adopting machine learning for classification model learning. Hu Yujin et al [2] proposed a Bayesian text classification method based on the vector space model. In this paper, the vector space model is first used to represent the text features, and then come to the Bayesian text classification method, which uses probability statistics to classify the documents. At the same time, the vector space can make the complexity of the problem greatly reduced. Li Aiping et al [3] refined the chapter-level sentiment analysis to vectors, used different analysis strategies for the sentiment analysis of different kinds of vectors, calculated the sentiment weights of various sentences, and finally synthesized the sentiment values of sentences into the sentiment tendency of the chapter according to a specific synthesis algorithm. Gamon et al [4] obtained the data of the depressed and non-depressed patients from Twitter, and according to the statistics of the data, the least squares method was used for the data classification. The least squares method was utilized to conduct regression analysis on the data. Statistical information about the time of users' posting of blog posts was analyzed to analyze the time difference between the two types of users' posting of blog posts; The degree of correlation between users' characteristics and depressive disorders was analyzed using the Pearson correlation coefficient method, etc. Wang X et al [5] used the sentiment analysis method to analyze the content of microblogs and assigned different weights to different words according to the words. Finally, the sentiment score of the sentence was calculated to determine the overall sentiment tendency of the sentence.

This paper focuses on the application of BiLSTM(Bi-directional Long Short-Term Memory) models and Attentional Mechanisms in the analysis of adolescent depression and compares the advantages and disadvantages of the two models.

2. BiLSTM model

2.1. LSTM

LSTM(Long Short-Term Memory) is a type of Recurrent Neural Network (RNN) that excels at modeling temporal data, including textual data, thanks to its unique design characteristics. There are several ways to combine word representations into sentence representations, such as summation (where the representations of all words are summed) or averaging. However, these methods overlook the positional order of words within a sentence. To capture dependencies over longer distances, LSTM models are more effective. This is because LSTM has the ability to learn which information should be retained and which should be discarded during the training process. Therefore, LSTM models excel at capturing the intricate relationships between words in a sentence.

2.2. The Reason to Use BiLSTM

BiLSTM, short for Bi-directional Long Short-Term Memory, is a combination of forward LSTM and backward LSTM. Both are commonly employed in natural language processing tasks to model contextual information. LSTM alone has a limitation when it comes to encoding back-to-front information in sentences. However, for more nuanced categorization tasks, such as differentiating between strong positive, weak positive, neutral, weak negative, and strong negative sentiments, it becomes crucial to consider the interactions between emotion words, degree words, and negative words. In such cases, BiLSTM excels at capturing bidirectional semantic dependencies, enabling a more accurate representation and understanding of the sentence.

2.3. Principle of operation of BiLSMT

LSTM is a type of recurrent neural network (RNN). In practical applications, traditional RNNs encounter challenges such as Gradient Vanishing, Gradient Explosion, and limited ability to incorporate long-range dependencies. To address these issues, LSTM was introduced. While LSTM shares a similar overall structure with RNN, its key improvement lies in the incorporation of three gates within the hidden layer, denoted by h. The formula for the LSTM unit is as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{2}$$

$$g_t = tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{cc}c_{t-1} + b_c) \tag{3}$$

$$c_t = i_t g_t + f_t c_{t-1} \tag{4}$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o) \tag{5}$$

$$h_t = o_t \, tanh(\, c_t) \tag{6}$$

BiLSTM consists of two basic LSTMs, compared with LSTM, BiLSTM fully considers the before and after contexts of sentences, extracts bi-directional semantic features, has greater predictive ability than LSTM models, and outperforms LSTM models in handling multivariate time series data [6,7]. BiLSTM is an LSTM in both directions, and ultimately the two directions are the results which are merged and output denotes forward output and $\overleftarrow{h_i}$ denotes backward output, and the calculation formula is:

$$h_i = [\overrightarrow{h_i} \oplus \overleftarrow{h_i}] \tag{7}$$

After extracting the text vectors using the pre-trained text model, the vectors are fed into the LSMT model and then normalized using the SoftMax function and finally, the probability values are calculated:

$$y = soft\,max(\, h_i) \tag{8}$$

3. Attention Mechanism

3.1. Attention

At its core, Attention is a mechanism that selectively filters crucial information from a vast amount of data and directs its focus towards this significant information, while disregarding the relatively less important content. The weight assigned to each piece of information determines its importance and dictates the level of emphasis placed on its corresponding Value, which represents the specific information being considered. This concept is visually illustrated in the figure provided below:

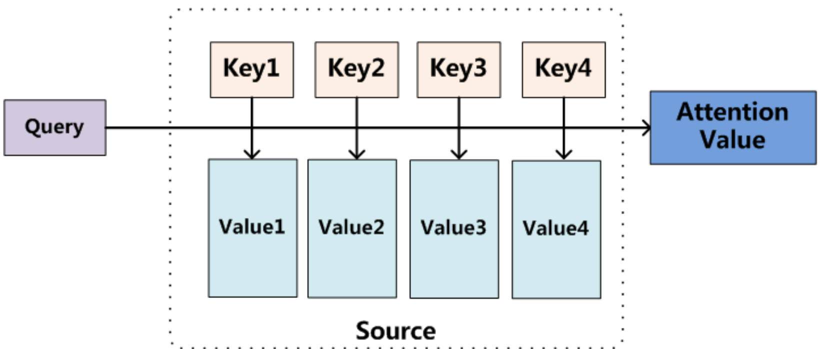

Figure 1. The principle of Attention Mechanism [8].

3.2. Principles of the Attention Mechanism

The specific computation process of the Attention Mechanism can be broadly summarized into two steps. The first step involves calculating weight coefficients based on Query and Key, while the second step entails weighting and summing the corresponding Values using these weight coefficients. The first step can be further divided into two stages: the initial stage focuses on determining the similarity or relevance between Query and Key, while the subsequent stage normalizes the obtained scores. This way, the computation process of Attention can be abstracted into three stages, as depicted in the accompanying figure.

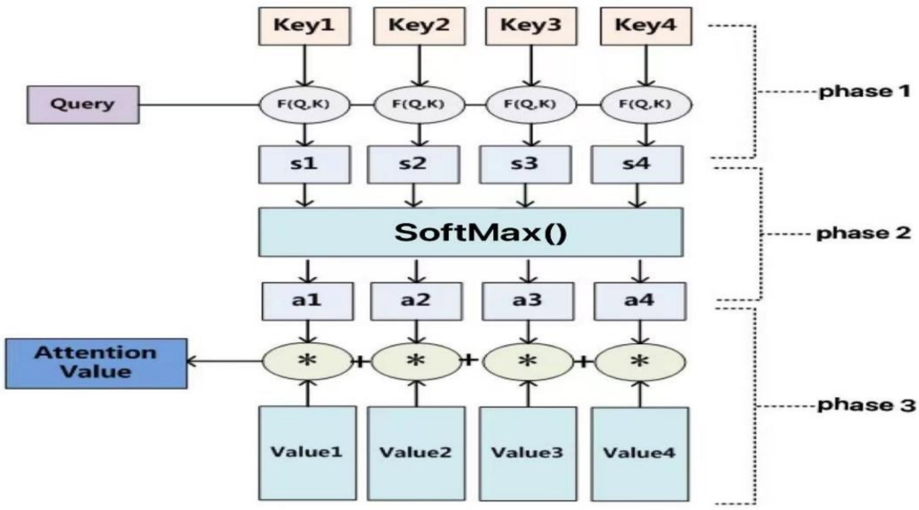

Figure 2. Three phases [8].

In the initial stage, various functions and computational mechanisms can be utilized to calculate the similarity or correlation between Query and Key. The commonly employed methods encompass finding the vector dot product of the two, determining the vector cosine similarity, or incorporating an additional Neural Network for computation. Thus, the following approaches can be employed:

$$Similarity(Query, Key_i) = \begin{cases} Query \cdot Key_i \\ \frac{Query \cdot Key_i}{\|Query\| \cdot \|Key_i\|} \\ MLP(Query, Key_i) \end{cases} \tag{9}$$

The scores produced in the first stage can vary depending on the specific method of generation. The second stage involves a computation method similar to SoftMax, which serves to numerically convert the scores generated in the initial stage. This process not only facilitates normalization of the raw scores but also arranges them into a probability distribution whereby the sum of the weights of all elements adds up to 1. Furthermore, this mechanism makes it possible to more strongly emphasize the significance of crucial elements. As such, the following formula is usually employed for this computation:

$$a_i = Softmax(Sim_i) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}} \tag{10}$$

The outcome of the second stage calculation, represented as a_i, corresponds to the weight coefficient assigned to $Value_i$. Subsequently, the Attention value can be derived by weighting and summing these values:

$$Attention(Query, Source) = \sum_{i=1}^{L_x} a_i \cdot Value_i \tag{11}$$

3.3. Diverse Attention Models

Self-Attention Mechanism [9] (SAM) is an Attention Mechanism for calculating the degree of correlation between each element and other elements in a sequence. In Self-Attention Mechanism, each element calculates its own correlation with other elements in the sequence and assigns weights based on these correlations. The Self-Attention Mechanism is commonly used to deal with the relationships between elements within a sequence, such as text generation, language translation, etc. It captures dependencies and contextual information between different elements.

Multi-Head Attention [10] (Multi-Head Attention) is a mechanism that expands on the Self-Attention Mechanism. It captures different levels and aspects of information by using multiple independent Self-

Attention Mechanisms simultaneously. Each Attention Head learns different correlations in the sequence and generates corresponding weight assignments. Multi-Head Attention Mechanism introduces multiple independent Attention Heads on top of the Self-Attention Mechanism to better capture different levels and aspects of information in a sequence. It is often used to handle more complex tasks such as semantic understanding and semantic matching.

Cross-Attention is an Attention Mechanism used to compute the correlation between two different sequences. In Cross-Attention, one sequence (e.g., question) is considered as a Query and another sequence (e.g., context) is considered as Key and Value. By calculating the correlation between the query keys and the keys, weights can be assigned to the values, thus enabling the modeling of the correlated information between the two sequences. The Cross-Attention Mechanism is commonly used to deal with the relationship between two different sequences, such as the association between questions and contexts in question-and-answer systems. It is able to interact the relevant information of the question with the context to better understand the context and generate accurate answers.

4. Advantages and Disadvantages of Bilstm and Attention Mechanism in Text Sentiment Prediction

4.1. Advantages of Attention

Effective Handling of Long Sequences: While BiLSTM is an improvement over RNN, it still encounters difficulties in effectively capturing information at the beginning of long sequences. On the other hand, the Self-Attention Mechanism is not constrained by the sequential order of words. Instead, it exploits similarity calculations to extract relevant information. As a result, it avoids any loss of information, even in lengthy sequences.

Independence from Temporal Information: Unlike BiLSTM, which relies on sequential calculations and necessitates considering the previous moment to compute the next one, the Self-Attention Mechanism is not bound by such temporal dependencies. It can parallelize computations because similarity calculations do not require strict sequential processing. This characteristic vastly enhances computational efficiency, allowing for faster processing of large-scale data.

Flexibility in Capturing Interdependencies: The Self-Attention Mechanism excels at capturing interdependencies between elements within a sequence. By calculating the similarity between words, it can assign higher weights to more important elements. This flexibility enables the model to focus on key information while effectively disregarding less relevant elements. Consequently, it enhances the model's ability to understand complex relationships and make more informed predictions.

Scalability and Adaptability: The Attention Mechanism has shown remarkable scalability across different tasks and domains. It adapts well to various types of input data, including text, images, and audio. Additionally, the Self-Attention Mechanism can be easily incorporated into Deep Learning architectures, allowing for seamless integration and enhancing the performance of models across diverse applications.

4.2. Disadvantages of Attention

Despite its numerous advantages, the Attention Mechanism has specific limitations that need to be taken into account. One such disadvantage is that it may not perform as well as BiLSTM when dealing with sparse or limited data. This is because the Attention Mechanism focuses on key information and ignores unimportant information, which may lead to inadequate training if there is a lack of important data points.

Another potential limitation of the Attention Mechanism is that it directs its focus towards a select subset of crucial information, without explicitly acquiring temporal information. In certain tasks that require contextual judgment, such as language modeling, temporal information is precisely important. The Attention Mechanism may not work as well as BiLSTM in such cases because it does not explicitly account for the sequential nature of the input.

Of course, the introduction of an Attention Mechanism on top of BiLSTM has shown better performance in sentiment analysis [11], and many experimental results have demonstrated that the

BiLSTM-Attention model outperforms each of the two models mentioned above, by enabling the model to focus on information-rich words and phrases.

5. Conclusion

In terms of text sentiment prediction, many scholars have conducted research and discussion, using different models for simulation experiments. The BiLSTM model and the Attention Mechanism mentioned in this paper are both very common models in Text sentiment analysis, and both are widely used and compared to capture and understand the contextual information in text. Meanwhile, BiLSTM and Attentional Mechanism are techniques that also complement each other in sentiment analysis tasks. BiLSTM captures serial dependencies and word order, while the Attentional Mechanism provides a mechanism to dynamically focus on relevant information, which in turn improves interpretability and performance. Perhaps this paper needs to find more experimental literature to verify the specific properties of different models so that the advantages of different models in textual sentiment prediction can be learned more intuitively. Meanwhile, combining these two techniques has shown good results in sentiment analysis tasks [12], and combining the different models with BiLSTM or Attentions in future research to make up for the disadvantages of the two models mentioned above can play a greater role in the detection of text depression.

References

[1] Liu B. Sentiment Analysis and Opinion Mining[C]// Synthesis Lectures on Human Language Technologies. Morgan & Claypool, 2012:167.
[2] Yujin Hu, Xiaoling Zhou, Ling Ling, Xuelin Wang. A Bayesian text classification method based on vector space model[J]. Computer and Digital Engineering,2004,(06):28-30+77.
[3] Li Aiping, Di Peng, Duan Liguo. Chapter sentiment analysis based on sentence sentiment weighting algorithm[J]. Small Microcomputer Systems,2015,36(10):2252-2256.
[4] Gamon M, Choudhury M D, Counts S, et al. Predicting Depression via Social Media[C]//AAAI. 2013.
[5] Wang X, Zhang C, Ji Y, et al. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network[J]. Lecture Notes in Computer Science, 2013, 7867:201-213.
[6] GRAVES A, JAITLY N, MOHAMED A. Hybrid speech recognition with deep bidirectional LSTM[C]//2013 IEEE Workshop on Automatic Speech Recognition and Understanding. olomouc IEEE Press, 2013:273-278.
[7] SIAMI-NAMINI S, TAVAKOLI N, NAMIN A S. A comparative analysis of forecasting financial time series using ARIMA, LSTM, and BiLSTM [DB/OL]. [2023-01-18]. http://arXiv:1911.09512v1
[8] Zi You Ji YI, Explanation of the Attention Mechanism in Deep Learning, Apr 3, 2023. https://zhuanlan.zhihu.com/p/379722366
[9] MICHEL P,LEVY O,NEUBIG Graham. are sixteen heads really better than one?[J]. arXiv preprint arXiv:1905.10650,2019.
[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach,California,USA :nips,2017:5998-6008.
[11] Zhi Linyun,Qu Jinshuai,Fan Jing et al. Text sentiment analysis based on BERT-BiLSTM-Attention[J/OL]. Journal of Yunnan University for Nationalities(Natural Science Edition):1-11[2023-08-24].http://kns.cnki.net/kcms/detail/53.1192.N.20230307.1708.006.html
[12] Z.Q. Xue,G.H. Zhang,M.H. Wang et al. Research on conversational text depression recognition based on Attention-BiLSTM model[J]. Computer Knowledge and Technology,2023,19(07):38-41.DOI:10.14004/j.cnki.ckt.2023.0477.