A Compression-based BiLSTM for Treating Teenagers’ Depression Chatbot

Jun-jie YIN
Guangdong Country Garden School, Guangzhou, China

Keywords: Chatbot, NLP, Compression-based BiLSTM, CNN, Deep learning.

Abstract. Anxiety is a psychological condition that often occurs during adolescence. Due to lack of relief and counseling, teenager’s psychological anxiety may gradually develop into anxiety. Chatbot can be used as a new tool to relieve anxiety among teenagers. However, the natural language understanding techniques currently applied to Chatbot still have problems, such as lack of effective data, high training complexity, and lack of interpretability of the network. This paper proposes a compression-based bidirectional Long Short-Term Memory depth neural network structure. The main objective is to reduce the complexity of the parameters further, and to make the network of each layer have certain interpretability by means of the reduction of sparsity. Under our own collection of teen depression text data, this structure shows a better performance than traditional networks.

Introduction

Anxiety is a mental health condition that often occurs during adolescence. Tensions of native family relationships, excessive narcissism, increasing materialism and consumerism have caused more than a quarter of teenagers to struggle for psychological anxiety at some point in their lives. About one in every 20 people will have severe anxiety disorders. Psychological anxiety begins with long-term accumulation of isolation, sensitivity, anxiety, sadness, etc., but more than 90% of anxious teenagers have never voluntarily confided or asked for help from teachers, friends or doctors because of concerns about negative impressions. Because of lack of relief and counseling, psychological anxiety, which may gradually develop into an anxiety disorder, leading to various family and social problems.

Compared with low-anxiety adolescents, high-anxiety adolescents not only show higher anxiety and stress, but also show higher anger, sadness and fatigue, as well as lower happiness and happiness. Compared with achievement-oriented pursuits, stronger diet and smoking desires, and more tobacco use, they reported fewer conversations and fewer recreational activities. Gender differences are small. Although tending to spend less time with peers, high-anxiety adolescents are more likely to show reduced anxiety with their friends [1].

Human-machine voice is gaining momentum as a computer interaction technology. Recently, the number of voice-based search engines and assistants, such as Siri, Google Chrome, and Cortana, has soared. Natural Language Processing (NLP) techniques (such as NLTK for Python) can be applied to analyze speech and can provide intelligent responses through the design engine to provide appropriate human response. This type of program is called Chatbot [2]. The Chatbot is a virtual human-computer interaction system derived from Chat Robot that allows users to participate in a conversational experience through text input, which can be used on web browsers, mobile devices and chat platforms. Technologies such as natural language processing and deep learning have made breakthroughs in recent years, making Chatbot, which simulates human conversations, widely available in call centers and customer services. The private virtual assistants are also usually in the form of Chatbot.

Designing and developing Chatbot as a new interface for providing health information to young people requires a new understanding of new user needs and motivations and privacy [3]. It is not straightforward to develop a successful session interface based on Chatbot to provide health information. Using a chat bot as a conversational user interface to younger users, and the way
communicate and communicate has changed dramatically. The paper [4] has used supervised data to train universal sentence representations instead of usual unsupervised data, which show a better performance. The method they developed can be well suited to other NLP tasks by showing the suitability of natural language inference.

Motivational Interviewing (MI) is a style of psychological counseling that focuses on changing the client's behavior, and is often used when the client is addicted to substances. Two NLP systems based on DSF and RNN are used in the study [5] to automatically code MI sessions, and predictions from these models are compared to human ratings from a large sample. The DSF model performs slightly better at utterance- and session-level agreement than the NLP model, as the RNN model is trained with more formal sources, but poor agreement is observed with both models in some cases. NLP models have the potential to allow clinical supervision to be a practice.

Word embeddings enable NLP models to capture relationships between words. The study compares word embeddings trained from different sources and qualitatively and quantitatively evaluates the embeddings. It is found that embeddings trained from clinical notes and biomedical publications better capture relationships than those trained from Wikipedia and news, but is not necessarily better at performance [6]. NLP systems can be built on a maximum entropy approach based on real-valued features and constraints built on these features. The literature [7] select the model with the greatest entropy within the constraints, which is the same as the exponential model which best predicts the sample of data. Then we select a set of features to include in the model. The algorithm is feasible for context-sensitive modeling.

Opinion mining, also known as sentiment analysis, is an essential approach to analyze data. In this paper, NLP techniques are reviewed for text preprocessing, opinion mining approaches [8] are investigated for different situations, and DL approaches for opinion mining are explored in detail. The paper [9] describes that NLP uses techniques in order to learn, understand and produce human language.

However, there are still some difficulties in the current NLP technology applied to Chatbot chat robots for treating depression in adolescents. Although the current natural language processing system mainly uses the latest deep learning technology, the current Chatbot used to treat adolescent depression still has some specific problems that are difficult to achieve, such as the lack of effective data in the actual scene, and the length of training. The complexity and complexity are difficult to reduce, and the neural network lacks interpretability and meaning. In this paper, a compression-based BLSTM deep neural network structure is proposed. The main purpose is to further reduce the parameter complexity of the deep learning network structure and make each layer network have certain interpretability. Our method exhibits better performance than the traditional network under the collection of adolescent depression text data collected by ourselves, and can maintain a certain response time and robustness with high input data volume.

The paper consists of the following: First, the mainstream deep learning natural language processing scheme is provided. Then we propose the Chabot system and method based on BLSTM structure, and compare and test various methods on the data set of adolescent depression dialogue recorded by ourselves. The test results are provided in the fourth part, and the fifth part is summary and future.

**NLP Methods by Deep Learning**

All NLP tasks can be thought of as tasks that assign labels to words. The traditional NLP approach is to extract a rich set of hand-designed features from a sentence and then import those features into a classic shallow classification algorithm, such as a support vector machine (SVM), usually with a linear kernel. The choice of features is a fully empirical process, mainly based on trial and error, and feature selection depends on the task, which means additional research on each new NLP task. Complex tasks like SRL then require a large number of potentially complex features, which makes such systems slow and difficult to handle for large-scale applications. Instead, we advocate a deep neural network (NN) architecture that trains in an end-to-end manner.
The input sentence is processed by several layers of feature extraction. The features in the deep layers of the network are automatically trained by backpropagation to be related to the task. A general deep architecture suitable for all NLP tasks [10] is given in Figure 1. The first layer extracts the characteristics of each word. The second layer extracts the feature from the sentence and treats it as a sequence with local and global structure (i.e., it is not treated like a bag of words). The following layers are classic NN layers.

**Figure 1.** The general deep NN architecture for NLP.

Time-Delay Neural Networks (TDNNs) are selected when modeling long-distance dependencies is required. The classical TDNN layer performs a convolution on a given sequence $x(\cdot)$ and outputs another sequence $o(\cdot)$ whose value at time $t$ is:

$$o(t) = \sum_{j=-d}^{d} L_j x_{t+j}$$

(1)

Where $L$ is the parameter of the layer trained by backpropagation.

**CNN-based Natural Language Processing**

Convolutional neural network architecture is one system, and when inputted with a sentence, its output are language processing predictions such as semantic roles. The entire network is trained by multitask learning. As an exception, for shared tasks, the language model learns by semi-supervised learning. This paper demonstrates how multitask learning and semi-supervised learning work together to improve the generalization of shared tasks. The purpose is to establish a NLP that learns relevant features of the current task with little prior knowledge. This can be applied to numerous NLP tasks. Bengio et al. proposed a neural language model that can learn the distributed representation of words [11], which is shown in Figure 2.

**Figure 2.** Representation of the Neural Language Model.

**RNN-based Natural Language Processing**

Recurrent neural network and convolutional neural networks are predominant in handling NLP tasks [12]. Alexis Conneau et al. have proposed a novel architecture by using deep layers that would typically see in computer vision to perform text processing, and have improved current classification tasks, which is shown in Figure 3. They have concluded that with more depth, the performance of the model will be enhanced. This is the first time deep convolutional networks are applied into NLP and offered an insight into how this can improve tasks related to NLP. From building a latent semantic analysis model that used the bottom-up technique in digital games to introducing and evaluating an interface like GameNet [5], which utilizes skills from Natural Language Processing to advance a
bottom-up approach on game studies. The research has shown that scholar can be more effective in studying the area of games that they are unfamiliar.

**LSTM-based Natural Language Processing**

The paper [13] recognizes the abundance of resources in radiological reporting, which may become a good indicator for clinical care and supporting research. Due to their forms in the free-text clinical narrative, information extraction is crucial to the process. Natural language processing techniques can be used to perform automatic identification and information extraction. Common Natural Language Generation (NLG) uses employ rules and heuristics, causing the responses generally to be monotonous and rigid. Wen and his peers have developed LSTM structure based generator [14]. Using cross entropy loss function, the LSTM based generator could learn from unaligned data and optimized the sentence planning, which is shown in Figure 4. This approach proved to have fewer heuristics and improved performance than the previous methods. This result corresponds further with subjective evaluation by judges.

![Figure 4. Semantic Controlled LSTM cell.](image)

The BLSTM architecture is defined as:

\[
\bar{h}_t = \tanh(W_h + h_t)
\]

(2)

\[
\alpha_i = \frac{e^{\tau_i}}{\sum e^{\tau_i}}
\]

(3)

\[
u = \sum \alpha_i h_i
\]

(4)

Where \(h\) are the output hidden vectors of a BiLSTM. The \(\alpha\) represent the score of similarity between the keys and a learned context query vector \(u\). These weights are used to produce the final representation \(u\), which is a weighted linear combination of the hidden vectors.

**Compression-based BLSTM Structure**

As social media grows in scale, the amount of hate speech, also known as hostile speech, profanity and cyberbullying increases. Supervised learning NLP techniques are explored in this essay [15] to automatically detect hate speech, and limitations of these techniques are examined. The possibility of constructing a benchmark dataset for hate speech recognition is also proposed. The Dynamic Memory Network (DMN) is a neural network architecture [16] that processes input sequences, creates episodic memories (memories relevent to the self) and produces relevant answers. This can be trained to thoroughly on many types of tasks, including sequence modeling for part-of-speech tagging (WSJ-PTB). Training techniques rely solely on trained word vector representations and input-question-answer triplets.

In this tree-based convolutional neural network (TBCNN) for programming language processing, a convolution kernel (image processing) is outlined over the program's abstract syntax trees to process structure information [17]. The Stanford CoreNLP toolkit [18] is a flexible pipeline that lays out core natural language analysis. Applications include in research NLP and users of open source NLP technology. Different from others, this system follows a simple, straightforward interfaces and high quality analysis components and does not use a large amount of associated but unnecessary baggage.
where $c$ is the weighted compression parameter, which further reconstruction of sparse coefficients by using compressed sensing theory. The $h$ are the output hidden vectors of a compressed BiLSTM. It can be seen from the formula that the bidirectional Long Short-Term Memory neural network structure based on compressed sensing can further adjust the parameter complexity of the past BLSTM structure; in addition, the sparsity of the compressed parameters is further reduced. To make each layer network have a certain degree of interpretability.

**Experiments Based on Teenager Depression Dataset**

To assess the performance of the above methods, we collected data on depression tendency screening data for adolescents (average 16 years, half of boys and girls) at the school. The survey was conducted using a personality scale for adolescents. The scale consisted of 22 items with a total of 3 subscales: personality, anxiety, and depression. The total score is divided into 25 items and can reflect the seriousness of students' bad emotions. It also includes three factor points, which correspond to: personality, anxiety tendency, and depression tendency. Anxiety factor: a total of 25 points, a total of five. Scores 1-5 points from no symptoms to severe symptoms, and a potential anxiety tendency if the total score exceeds 60%. Depression factor: A total of 42 points, a total of 9. Each score is between 1 and 5 points, and if the total score exceeds 60%, it can be considered a potential depression. Personality factor: A total of 24 points, a total of 12. Each item has a maximum of 2 points and a minimum of 1 point. A score of 2 in one item indicates that the character is extroverted and one point is introverted.

![Figure 5. Average accuracy with BiLSTM and Compressed BiLSTM.](image)

Based on this scale, text dialogue data is designed. The problem covers a lot of life and psychological content of teenagers. For example, whether I am easy to fall asleep and sleep well overnight/no or very little time/small part of time/quite a lot of time/Most or all of the time? There are 34 similar problems, ranging from one to five. Less than 20 is normal, and more than 30 is an anxiety. More than 60% of the scores are subject to anxiety. In order to effectively evaluate the proposed method, the above data is designed to be segmented into: Task1: 80% training, 10% verification, 10% test. Task2: 70% training, 20% verification, 10% test. Task3: 60% training, 20% verification, 20% test. Task4: 50% training, 30% verification, 20% test. Task5: 40% training, 20% verification, 30% test. Moreover, compare the performance of BiLSTM and compressed BiLSTM network structures in these tasks.

According to the above figure, the compressed BiLSTM error rate proposed in this paper is lower than that of the general BiLSTM model. Because our initial data was less than a thousand copies, the best effect on the original compressed BiLSTM was 78% accuracy. Then we try to use the sigmoid activation function to give a series of conversations a probability, between 0-0.5 is negative, 0.5-1 is positive. In addition, use k-fold to verify the accuracy of the model, while further increasing the sample to more than 1600. The final correct rate is 85% or more.
Conclusion

This paper proposes a compression-based BLSTM neural network structure. While further reducing the complexity of the parameters, the sparseness of the data also makes the network of each layer, have a certain degree of interpretability. Compressed BLSTM exhibits better performance than traditional networks under the actual collection of teen depression text data.

References


