International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

WORD BASED LANGUAGE MODEL USING LONG SHORT TERM MEMORY FOR DISABILITIES

[1]Guruvammal.S, [2]Dr.T.Chellatamilan, [3]Dr.L.Jegatha Deborah

ABSTRACT—*Identification of Autism Spectrum Disorder (ASD) among children plays a vital role to develop their social interaction through early intervention programme. Our proposed model categorizes the children who affected with ASD from typically developed children based on their linguistic knowledge. Existing methods are lagging with short-term memory and early saturation of learning rate. Our proposed model solves the demerits of existing approaches and has the advantages of classifying the children with ASD accurately with high precision and recall as well as a new model of LSTM based deep learning neural network has used. Word prediction is useful for disabled persons and IQ of Autism children will be tested based on latent semantic analysis. The experimental results showed in this paper promisingly.*

Keywords*-- Autism Spectrum Disorder, Long Short-Term Memory, Language Modeling, Natural Language Processing*

I. INTRODUCTION

Modeling language is a subtask of natural language issues. Natural Language Processing is a field where computer science, artificial intelligence and linguistics converge. NLP is an area of machine learning capable of understanding, analyzing, manipulating and potentially producing human language. A language template is a distribution of probability over word series. Language Models are used in various applications[4] such as voice recognition, machine translation, automated description, emotion analysis and text classification.

The probability of the next word in the sequence can be predicted by a language model based on the words already detected in the sequence. This model is called a word-based modeling of language. Deep learning approaches for certain specific language problems produce the most modern results [1]. A single model of deep learning can learn a word's meaning and perform language tasks. Deep neural networks usually have a lot of hidden layers within them and are generally recurrent. Recurrent Neural Network is composed of Long Short Term Memory network units is often called an LSTM network.

LSTMs are designed to avoid the long-term dependency problem. Its default behavior is remembering information for long periods. LSTM was introduced by Hochreiter & Jurgan Schmiduber [2]. It is used for speech recognition, language modeling, sentiment analysis and text prediction.

[1] *Computer Science and Engineering,University College of Engineering,Tindivanam, India, guruiragulraj@gmail.com*

[2] *Information Technology,SITE, VIT,Vellore, India,chellatamilan@gmail.com*

[3] *Computer Science and Engineering,University College of Engineering,Tindivanam, India,blessedjeny@gmail.com*

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

LSTM has contacts with feedback. It can not process single data points and can process entire data sequences. The cell of the LSTM unit retains values over arbitrary periods of time. LSTM treated gradient problems that exploded and vanished.

Autism-related Natural Language Processing (NLP) technologies are (1) picture-based emotion detection system (2). Improved Alternative Communication (AAC) deficiency is resolved by the NLP. Many patients with Autism tend to improve their language-based skills through highly interactive programs. NLP loads it with full characteristics.

NLP provides high features to help people with communication difficulties develop their language and gain consistent performance information of their respective inputs. NLP solutions to AAC systems are highly efficient and mandatory for the removal of people's communication disorders [9].

II. RELATED WORKS

A conventional neural network is unable to learn from the previous events because the information does not pass from one step to the next. But Recurrent Neural Network learns information from immediate previous step. When train the RNN, the information goes in loop again and again, so large number of updates to neural network weights. It is happened due to the accumulation of error gradients during an update and it results in an unstable network.

To overcome this drawback of RNN, new variant of RNN model i.e. Long Short Term Memory (LSTM) developed. LSTM unit (Figure 1) uses gates to control the memorizing process.

Information passes through many LSTM units. Each LSTM unit has three components to memorize the information. They are:

- Cell state
- Forget gate
- Input gate
- Output gate

All three gates regulate information flows into and out of the cell. The cell state keeps track of the dependencies in the input sequence between the elements. A new value flows into the cell through the input gate. The forgotten gate controls the cell's value. For measure the values, the output gate is used. One unit's output passes into the next unit and it transfers the data. We want the empty word in the text to be expected. For example:

“Shankar, who lives in India at the age of 37. He has a friend of Mala's mother. Mala works as a cook in a renowned restaurant in Chennai that he recently met in a class. Mala told him that she always had a passion for_”.

The network will learn to predict ' cooking ' from the dependence ' cook. ' There is a difference between what we want to predict about the data and where we want it to be predicted. This is referred to as long-term dependency’. When we work on above trigram, we need to consider long term dependency.

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

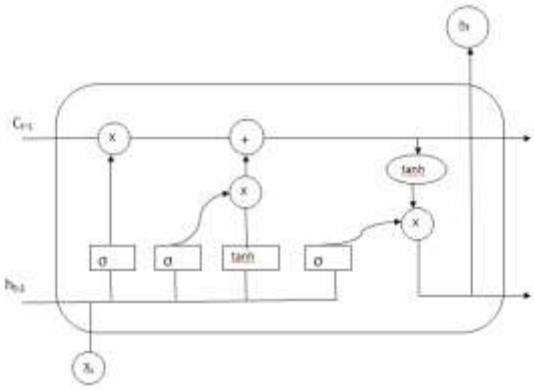

Figure 1: LSTM Unit

The LSTM unit has:

X: Scaling of Information

+ : Adding Information

σ : Sigmoid layer

tanh : tanh layer

h(t-1) : output of last LSTM unit

C(t-1) : Memory from last LSTM unit

X(t) : Current input

C(t) : new updated memory

h(t) : Current output

LSTM causes needless data to be overlooked. In the example given: When the feedback is ' He has a female friend Mala, ' Shankar's gender may be overlooked as the topic has shifted to Mala. Forget gate accomplishes this process. When new entries are received in the LSTM unit, the information will be stored in the cell state. Whenever the input is: ' Mala works as a cook in a famous Chennai restaurant that he recently met in a school alumni meeting, ' the words ' famous, ' ' school alumni meeting ' can be ignored and words like ' cook, ' ' restaurant ' and ' Chennai ' will be updated.

A sigmoid layer determines what parts of the cell state we're going to remove to determine the output. The proposed model recognizes that it is a noun linked to "cook" from its memory to predict the blank word; the result is "cooking". The model proposed learned from long-term dependency rather than immediate dependence.

Word forecasting is very useful for people who are slow in typing or lack a strong vocabulary[3]. There was a growing number of people with disabilities. Prediction tool is used to reduce the user's keystrokes and increase the speed of communication. Prediction refers to suggesting the typing of words and phrases.

The different algorithms [3] used in word prediction tool are categorized as:

- N-gram model
- Frequency-based model
- Recency-based model

For word prediction, Shashi Pal Singh et al[3] used a trigram model. The training corpus is divided into three-word chunks for the trigram algorithm and the frequency of the triplets is calculated and stored in the database.

The activity of prediction depends entirely on the corpus and the type of the corpus[5]. Deepa Nagalavi et al[6] develop a model in a newspaper page to identify individual items. The model predicts the preceding word with the likelihood of a word series knowing the preceding material. Based on previous words of n-1, the methodology of nth word prediction is called the hypothesis of Markov or the prototype of Markov. N-gram model follows the properties of Markov assumption.

In the name of Contextual LSTM, Shalini Ghosh et al.[7] designed a model for word prediction, next sentence selection, and the prediction of sentence topic. The model takes words encoded in 1- hot encoding from the input. And it is converted to an embedding vector and consumes word vectors one at a time.

Sanchari Sen et al. [8] explained approximate computing techniques to accelerate the execution of LSTMs. The researchers explained that how to reduce execution time and energy with minimal effect on the quality of outputs.

III. METHODOLOGY

The process of word prediction has three steps. They are:

- Data preparation
- Train language model
- Use language model

The data preparation process depends on how we intend to model it and how we intend to use it. We are designing a language model in this paper to generate new text sequences. The length of the input will be long for the words to predict to learn the context. We have developed a n-gram template in which the raw text is converted into sequences of 50 words input into one word output. Template N-gram is a statistical model to predict the next word. N-gram models estimate any word's probability based on previous n-1 words ' probability.

We need to turn the raw text into a sequence of tokens or words as the origin to train the template after loading the text into the memory. To clean the text, replace' -'with a white space, delete all punctuation from words to reduce the size of the vocabulary and normalize all words to reduce the size of the vocabulary. The result of a smaller vocabulary size is a smaller model that trains more quickly. The cleaned tokens can be shaped into sequences of 50 input words and 1 output word, i.e. 51 word sequences. For later loading, we can save the sequences to a new file.

Based on their context, LSTM is used to predict words. To learn the representation of words, an embedding layer is used. It expects entry sequences to consist of entries. We can map every term to a specific integer in our vocabulary and encode the sequences of inputs. We can convert the prediction to numbers when we make predictions and look up their related terms in the same mapping. In the Keras API, we used tokenizer class to encode all the unique words in the data and assign a unique integer to each of them.

We need to separate the words with array slicing into input and output elements. After separation, we need to have one hot encoding of the output word, i.e. converting the output word from an integer to a vector of 0 values. The model is learning to predict the distribution of probability for the next word. We need to specify to the embedding layer the vocabulary size and length of the input sequences.

We used 2 LSTM hidden layers with 100 memory cells each. More memory cells and a deeper network may achieve better results. To interpret the features extracted from the sequence, a dense fully connected layer with 100 neurons connects to the hidden layers of the LSTM. As a single vector, the output layer predicts the next word. To ensure that the inputs have the properties of normalized probabilities, a softmax activation function is used.

The model is assembled to match the equation with the categorical cross entropy loss. The model is learning a classification of multiple classes and this is suitable for problem with term prediction. The model will predict by naming predict category which returns the word index with the highest likelihood. At the end of each batch update, we can see a description of the results, including loss and accuracy.

In N-gram language model, word prediction is calculated by probability function using Bayes theorem.

$$p(w) = \prod_{i=1}^{n} p(w_i \mid w_{1,\dots,}w_{i-1})$$

N-gram model is expressed by:

$$p(w_i \mid w_{1,\dots,}w_{i-1}) \approx p(w_i \mid w_{\,i-n+1,\dots,}w_{i-1})$$

when n=1/n=2/n=3

IV. CONCLUSIONS AND FUTURE WORK

The children with Autism Spectrum Disorder (ASD) differ in IQ level. 31% of ASD children have intellectual disability i.e. Intelligent Quotient(IQ) < 70, 25 % are in the borderline range i.e. IQ is between 71 and 85 and 44% children have IQ scores in the average to above average range i.e. IQ >85. By giving constant practice on word prediction, language learning capacity will be improved in disabled children. As a future work, we have considered to improve language learning level of the disabled children through the auto-correct of their sentence i.e. one of the application of language modeling. Further research work will transform the undefined and meaningless information to clear and meaningful information.

REFERENCES

1. Piotr Klosowki, "Deep Learning for Natural Language Processing and Language Modeling", IEEE, 2018.
2. Nimesh Sinha, "Understanding LSTM and its quick implementation in Keras for sentiment analysis", Towardsdatascience, Feb. 20, 2018.
3. Shashi Pal Singh, Ajai Kumar, Daya Chand Mandal, Yasha Jadwani, "Word and Phrase Prediction Tool for English and Hindi Language", ICEEOT, 2016.
4. Nebojsa D. Grujic, Vladimir M. Milovanovic, "Natural Language Processing for Associative Word Predictions", IEEE, 2019.
5. M.P.Bhuyan, S.K.Sharma, "Effects of Prediction Length on Accuracy in Automatic Assamese Word Prediction", IEEE, 2019.
6. Deepa Nagalavi, M.Hanumanthappa, "N-gram Word Prediction Language Models to Identify the Sequence of Article Blocks in E – Newspapers", IEEE, 2016.
7. Shalini Ghosh et al., "Contextual LSTM (CLSTM) models for Large Scale NLP tasks", ACM, 2016.
8. Sanchari Sen, Anand Raghunathan, "Approximate Computing for Long Short Term Memory(LSTM) Neural Networks", IEEE, 2018.
9. Aravind Kannan, Dr.V.Thiruvalluvan, "Application of Natural Language Processing Techniques to Augmentative Communication Systems".