Advances and Applications in Mathematical Sciences
Volume 20, Issue 3, January 2021, Pages 377-382

TEXT SEQUENCE PREDICTION USING RECURRENT NEURAL NETWORK

AKASH KHARE, ANJALI GUPTA, ANIRUDH MITTAL and AMRITA JYOTI

ABES Engineering College

Ghaziabad, Uttar Pradesh, India

Abstract

This paper tries to show how long-term memory-recursive neural networks can be used to generate text sequences in real-time conditions, by predicting a single data point at a time. Next word prediction is an intensive problem in the field of NLP (Natural language processing). Word prediction is the problem of calculating which words are likely to carry forward a given primary text piece. The resulting system is capable of generating the next real-time word in a wide variety of styles.

1. Introduction

Next Sequence prediction differs from other types of supervised learning problems. The sequence applies an order to the observations that must be preserved when making training models and predictions. RNNs are trained for sequence data through processing and predicting the actual data sequences one at a time. The standard Recurrent Neural Network architecture, is still constructive, but it is not perfectly suited for such tasks. That's why we introduced a new RNN version that is Long Short Term Memory (LSTM), with multiplier (or "gated") connections that permits to determine the current technical character. Next Transition matrix is determined by the next character from one hidden state vector.

1.1. Objective

The aim of our work is to identify the sequence of words that are time-

2010 Mathematics Subject Classification: 60G25, 62M45, 92B20.

Keywords: Convolutional Neural Network, Deep Learning, Long Short-term memory (LSTM), Machine Learning, Next word prediction, Recurrent Neural Network, Natural language processing.

Received May 20, 2020; Accepted July 31, 2020

dependent and work on a real-time scenario namely predicting the next sequence of words to achieve efficiency and accuracy in writing text-related tasks. We will first read the previous work done in this domain and then try to implement those tasks in our project.

1.2. Problem Identification and Definition

Man does not consider new ideas in his each second. When going through the text, we accept every word on the basis of our interpretation of the preceding words. We've the ability to understand and respond i.e. our thoughts tend to be ceaseless. Typical neural networks have a major drawback that they shrink during such assumptions. For example, assume that we aim to guess what type of event is happening at each point in the novel. It is fuzzy to think that past events in the story can be used to inform later ones by the use of logic of any typical neural network. Recursive neural networks provided a solution to this problem.

2. Related Work

The Barman and Boora model digs into a data set containing transcribed Assamese words and predicts the next word in Assamese text, with an accuracy of 88.20% with the help of LSTM. His method is more difficult in different languages than English. The biggest problem in prediction of the next word in any of the local language is that our systems only recognize ASCII values which happens to be the mostly used format of text files in a computer system and internet where every alphabet, number, or special character is denoted with a binary number in a 7-bit representation.

Felix et. Al (1999) [1] solved the previously unacceptable tasks by the RNN using LSTM. Gates of forgetting were proposed to solve recurrent versions of the above mentioned problems. A language model based on, simple iterative neural network was presented by Mikolov et. Al (2010) [3] to make a prediction of the next word in ordinal data easier. In alternative work, Alex (2013) [4] talked about the utilization of LSTM to create complicated long-range ordered format.

A model to filter spam was proposed by Jindal and Liu [2] and they categorized the spam into different classes on product advertising blogs.

Delany et al [1] hand over a study of viable functions of filtering spam SMS, which relied on the ML approach compared to DL. Rafiq et al [2] suggested the SLEE framework (structural algorithm in ambiguous environment) to detect real-time spam. (0.93 precision). Popovac et al. [2] proposed a CNN-based architecture with a layer of pooling and convolution to filter SMS spam and 98.4% of accuracy was achieved. Jain et al [2] who used the LSTM network to filter SMS spam achieved an accuracy of 99.01%.

David Balderas (2018) [6] proposed using a combination of ANNs. Each of the three ANNs serves a different purpose. Deep neural networks (DNNs), which use internal connections to form predictive maps. LSTMs simulate a kind of memory with the help of recurrent lifter. Complex data can be decomposed for simpler analysis through layers with the help of Sensory neural networks (CNN). A combination of them can allow a machine or system to identify, recall, and make predictions according to input and past inputs. The main purpose of this article was to evaluate how the combination of ANNs in these new algorithms improve image sequence prediction.

(i) A combination of LSTM and CNNs to form CLNN and

(ii) Combination of CNN, LSTM and DNN to produce CLDN.

Alex Graves (2014) [1] in his paper showed the potential of long-term, short-term memory recurrent neural networks to produce both distinct and real-valued sequences with complicated, long-range order using next-stage prediction.

All related work leads to the fact that this sequence generation is best possible using multiindexed LSTM neural networks.

3. Proposed Methodology

We will use LSTM-a RNN (Recurrent Neural Network) to predict our sequence.

3.1. Recurrent Neural Network [RNN]:

Recurrent Neural Networks i.e. RNNs define a recurrence relation over time steps which is given by

$$S_t = f(S_{t-1} * W_{rec} + X_t * W_x). \tag{1}$$

Equation (1) defines an iterative relationship over time steps where step t at st time, state t is the extrinsic input at time, Wrec and Wx are weight .Memory to the model is given by the feedback loop because it can recall the data betwixt time steps.

The current state St can be calculated by the RNN from the previous state S_{t+1} and current input X_t or predict the current state from S_{t+1} to the current St and current Input X_t. Regrettably, we will ask the model to predict the succeeding words or the next word by passing an input of 40 characters. We'll add the new character and discard the first one and predict again. This will go on until we complete one whole word.

3.2. Long Short Term Memory [LSTM] :

Two major problems plague the RNN-extinction and explosion. In typical RNNs the multiplication of gradient signal with a weight matrix to a large number can be done. Thus, the vital role is played by the magnitude of the load of transition matrix.

The gradient signal becomes smaller at every training stage when the weight in matrix is small, thus slowing down learning too much or stopping it altogether. This is known as Vanishing Gradient. We can simulate the effect of the vanishing gradient by applying the sigmoid function several times.

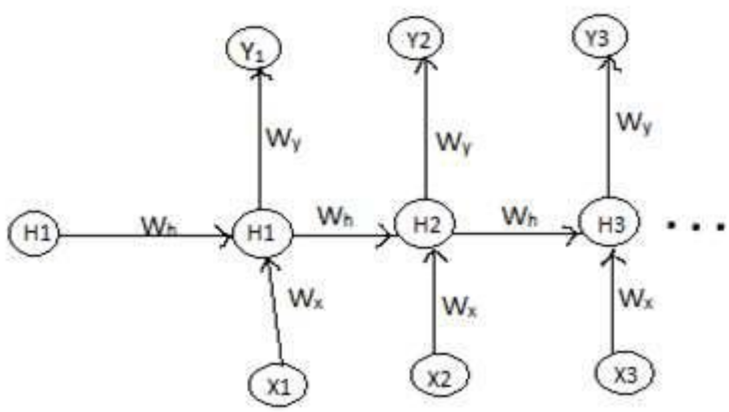

Figure 1. A Recurrent Neural Network [7].

In contrast, learning deviations can be caused by the large load in the matrix, to which the burst gradient points to.

LSTM is a unique type of RNN model. It learns long term dependencies. It introduces the new configuration - a memory cell made up of 4 elements : a neuron (that connects to itself), an forget gate, input gate and output gate.

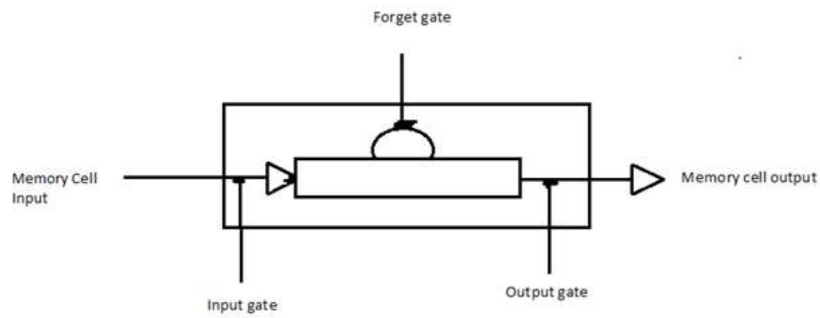

Figure 2. LSTM (gated) memory cell [7].

LSTMs fight a slowly fading difficulty by remembering the fault that can be propagated through layers and time. They allow long-term dependency learning by maintaining constant error. On the other hand, the explosion is controlled with gradient clipping. What it does is that, the gradient clipping doesn't allow the gradient to go over a predefined value.

4. Implementation

We've taken Project Gutenberg's Alice's Adventures in Wonderland, a project by Lewis Carroll. As our dataset. It will be used to feed the network and train it for the generation of text sequences. Smallest value of loss is not removed, all the other checkpoint is deleted from the directory. For example, when we run this example, "weight-correction-20-2.0818.hdf5" is the smallest loss I had gained.

A layer of LSTM is defined which is hidden with 256 memory units. The dropouts of probability 20 is used by Network. Softmax activation function is used to calculate the probability prediction of all the 47 Characters between zero and one, outer layer is thick layer. Ability prediction for each of the 47 characters between 0 and 1.

The network is not quick to train (approximately per epoch it takes 300 seconds). Model Check pointing is used to record weight every time a loss is observed to overcome the slowness and requirement of our optimization.

5. Results and Discussion

We have noted some examinations about the created text.

• This usually corresponds to the line format seen in the authenticated text of not greater than 80 characters preceding a new line.

• Characters are divided into clusters of word type and the majority of sets contain actual English words (such as “”, “less” and “was”), but many do not (such as “quiet”, “Pi” and “cadle”).

• Some words in the series make perception (e.g. “and the small toie”), but most of them do not (e.g. “no mistake!”).

It is very magnificent way of the character based prototype of this book. This gives you a perception of the training potential of the LSTM network.

The outcome are not correct. After developing more thick layers LSTM, we can see the increase in the standard of outcome.

References

[1] Alex Graves, Generating Sequences with Recurrent Neural Networks, (2014).

[2] Pradeep Kumar Roy, Jyoti Prakash Singh and Snehashish Banerjee, Deep Learning to Filter SMS Spam. (2019).

[3] Gend Lal Prajapati and Rekha Saha, REEDS : Relevance and Enhanced Entropy based Dempster Shafer approach for Next Word Prediction using Language Model (2018).

[4] Partha Pratim Barman and Abhijit Boruah (ICACC-2018), RNN based Approach for Next Word Prediction in Assamese Phonetic Transcription.

[5] IIya Sutskever, James Martins and Geoffrey Hinton, Generating Text with Recurrent Neural Network.

[6] David Balderas, Convolutional Long Short Term Memory Deep Neural Networks for Image Sequence Prediction (2018).

[7] Image source: deeplearning4j.org

[8] Tutorial by Andrej Karpathy titled, The Unreasonable Effectiveness of Recurrent Neural Networks.