A Deep Learning Approach for Mnemonic Generation

Vikas N S, Ajay M

Department of Information Science and Engineering,
M S Ramaiah Institute of Technology

Abstract

Mnemonics have been proved to be effective techniques as an aid for information retention [1][2][3]. The most common type of mnemonic phrase is a sentence that has the same initials as the list of words to be remembered. The words chosen here will provide the impression of a coherent overall meaning to the sentence, even if the sentence has no context. This paper showcases the use of a deep learning neural network, called the LSTM or Long Short-Term Memory, in conjunction with/without a bigram model, to provide such mnemonic phrases that aid the user in remembering his/her choice of words. The results were not significant enough to warrant a definite conclusion on the performance of the model. However, our observation was that there was a marginal improvement in the system's performance on the addition of a bigram model to the network.

Keywords: Deep Learning, LSTM, Bigram, Mnemonics

Architecture

Without Bigrams

The model comprises of a recurrent neural network using LSTM units. The input to the model is a matrix of the numerical encoded two-letter forms, scraped off the words of a meaningful sentence. In addition to the letters of the English alphabet, characters to handle the beginning and the end of lines are also part of the input. The model's first layer is the Keras Embedding Layer, which is initialized with random weights and will learn an embedding for all of the words in the training dataset. This is achieved through an unsupervised learning algorithm known as GloVe [4]. GloVe provides us with a suite of pre-trained word embeddings. This will then be used to set the weights in the Keras Embedding Layer, which in itself is used for the words in the training dataset.

Further, we create a matrix of one embedding for each word in the training dataset. We can do that by enumerating all unique words in the input list of words and locating the embedding weight vector from the loaded GloVe embedding.

The network further contains bidirectional LSTM layers that address the long-term dependency associated with the words of a sentence [5]. This is required for the information persistence to obtain words that form a coherent sentence. This is depicted in *Figure 1,* listed under the *Figures* section.

During the initial approach to the solution, the model's input consisted of the first two letters taken off the words from a sentence, which in turn was picked from the corpus of over 55,000 songs[1]. The letters themselves were encoded to integer representations before being fed to the first layer of the network.

The outputs of the model are dense vector representations for words that match the first two letters of the input words. The outputs of the model are measured against the labels – the sentence which was picked from a song. In this way, two letters off a word would provide a word, and a list of such words would produce a meaningful sentence.

The problem associated with the above approach was the incoherent meaning in the output words. This aberration was a consequence of the input consisting of words belonging to meaningful sentences in the first place. The words were part of a sentence and the sentence itself was part of a song. When the model was input with words that did not make up a meaningful sentence, it spewed out phrases that did not make sense.

---

[1] https://www.kaggle.com/mousehead/songlyrics

With Bigrams

Note: [Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style, available on the Home tab, in the Styles gallery, to get the spacing between table and note. Tables in APA format can use single or 1.5 line spacing. Include a heading for every row and column, even if the content seems obvious. A table style has been setup for this template that fits APA guidelines. To insert a table, on the Insert tab, click Table.]

Figures

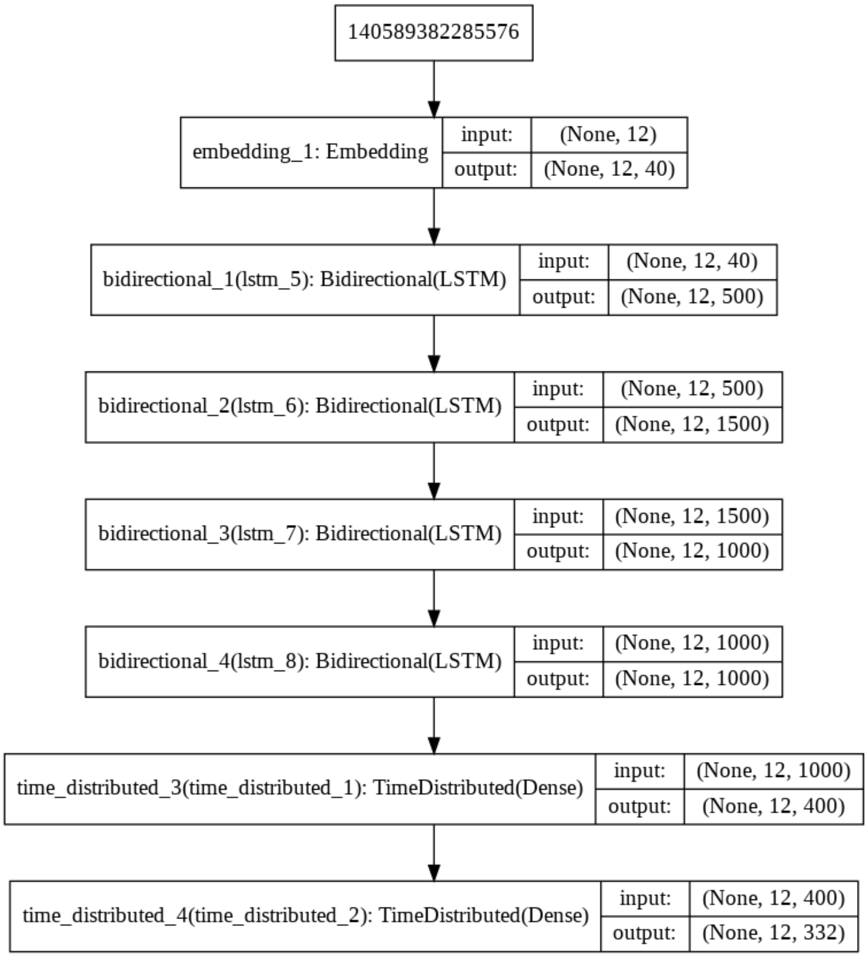

Figure 1. The architecture of the model used.

Discussion and Future Work

We have presented a model for mnemonic generation using LSTM units within an RNN, in conjunction with a bigram model. We believe that with the current style of pedagogy, there will always be a scope for topics that require the aid of information retention techniques. It is with that belief that we have strived to create such a system.

An apparent weakness of the model presented in this paper lies in not learning the rules of the grammar, instead focusing on the probability of the occurrence of a word based on the words surrounding it. Thus, we believe that a model which can incorporate the grammar rules of a language, alongside the techniques that we have presented in this paper, can significantly improve the results.

Related Work

1. Mountstephens, J. (2013, August). Mnemonic phrase generation using genetic algorithms and natural language processing. In *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 527-530). IEEE.
2. Korzeniowski, M., & Mazurkiewicz, J. (2017, June). Data-Driven Polish Poetry Generator. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 489-497). Springer, Cham.
3. The websites -
    i. http://www.mnemonicgenerator.com
    ii. https://spacefem.com/mnemonics

References

1. Roediger, H. L. (1980). The effectiveness of four mnemonics in ordering recall. Journal of Experimental Psychology: Human Learning and Memory, 6(5), 558-567
2. Megan Mocko, Lawrence M. Lesser, Amy E. Wagler & Wendy S. Francis (2017) Assessing Effectiveness of Mnemonics for Tertiary Students in a Hybrid Introductory Statistics Course, Journal of Statistics Education, 25:1, 2-11, DOI: 10.1080/10691898.2017.1294879
3. Rohwer W.D., Thomas J.W. (1987) The Role of Mnemonic Strategies in Study Effectiveness. In: McDaniel M.A., Pressley M. (eds) Imagery and Related Mnemonic Processes. Springer, New York, NY
4. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. [pdf] [bib]
5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.