

# Long Short-Term Memory Networks: Overcoming Vanishing Gradient Problem in Recurrent Neural Networks

James Crooks\*

Department of Systems Engineering and Automation, University of Seville, Seville, Spain

## Introduction

Long Short-Term Memory (LSTM) networks are a type of Artificial Neural Network (ANN) that are particularly well-suited to tasks involving sequence data, such as speech recognition, natural language processing and time series prediction. LSTM networks are a variation of the traditional Recurrent Neural Network (RNN) architecture, which is designed to deal with sequence data by processing it one element at a time while also maintaining a memory of previous inputs. The problem with traditional RNNs is that they can struggle with long sequences, particularly when the sequence has long-term dependencies or contains irrelevant information. When processing a sequence, an RNN stores a hidden state that is used to process the next element in the sequence. However, this hidden state can quickly become overwhelmed with information, leading to what is known as the "vanishing gradient" problem, where the gradients used to train the network become very small or even zero, making it difficult to learn long-term dependencies [1].

## Description

LSTM networks were introduced as a way of addressing this problem. Rather than a single hidden state, an LSTM network maintains a "cell state" that can be updated, reset, or accessed by "gates," which are layers that control the flow of information. These gates consist of sigmoid layers that can output values between 0 and 1, which are used to control how much information is allowed to pass through the gate. There are three types of gates in an LSTM network: the input gate, the forget gate and the output gate. The input gate determines how much of the new input should be added to the cell state, while the forget gate determines how much of the previous cell state should be retained. The output gate determines how much of the cell state should be outputted to the next layer in the network.

Another benefit of LSTM networks is that they can be stacked to create deep architectures that can learn more complex representations of the input sequence. Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) that is designed to overcome the problem of vanishing gradients in traditional RNNs. It was introduced by Hochreiter and Schmidhuber in 1997 and has since become a widely used model for various tasks such as speech recognition, natural language processing and time series prediction. The problem with traditional RNNs is that they suffer from the vanishing gradient problem, which means that the gradients used to update the weights during training become very small as they propagate backward through time. This problem makes it difficult for RNNs to capture long-term dependencies in sequences. LSTM networks overcome this problem by using memory cells that can store information over a long period of time. An LSTM network consists of memory

*\*Address for Correspondence: James Crooks, Department of Systems Engineering and Automation, University of Seville, Seville, Spain; E-mail: CrooksJames9@gmail.com*

*Copyright: © 2023 Crooks J. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

**Received:** 30 April, 2023, Manuscript No. sndc-23-96032; **Editor Assigned:** 02 May, 2023, Pre QC No. P-96032; **Reviewed:** 15 May, 2023, QC No. Q-96032; **Revised:** 22 May, 2023, Manuscript No. R-96032; **Published:** 30 May, 2023, DOI: 10.37421/2090-4886.2023.12.212

cells, input gates, forget gates and output gates. The memory cells are used to store information over a long period of time, while the gates control the flow of information into and out of the cell [2,3].

The input gate controls the flow of information from the input to the memory cell. It takes as input the current input and the output of the previous time step and outputs a value between 0 and 1, which represents the amount of information that should be allowed into the memory cell. If the gate output is 0, no information is allowed into the memory cell, while if it is 1, all information is allowed in. The forget gate controls the flow of information from the previous memory cell to the current memory cell. It takes as input the output of the previous time step and the current input and outputs a value between 0 and 1, which represents the amount of information that should be retained in the memory cell. If the gate output is 0, all information is forgotten, while if it is 1, all information is retained. The output gate controls the flow of information from the memory cell to the output. It takes as input the output of the previous time step and the current input and outputs a value between 0 and 1, which represents the amount of information that should be output. If the gate output is 0, no information is output, while if it is 1, all information is output.

The memory cell itself is a state vector that stores information over a long period of time. It is updated using the input and forget gates, which control the amount of new and old information that should be retained in the cell. The LSTM network can be trained using Back Propagation Through Time (BPTT), which involves calculating the gradients of the loss function with respect to the weights at each time step and propagating them backward through time. The gradients are then used to update the weights using an optimization algorithm such as Stochastic Gradient Descent (SGD). One advantage of LSTM networks is that they can handle input sequences of variable length. This is because the gates control the flow of information, so the network can learn to ignore irrelevant inputs and focus on the important ones. Another advantage is that they can capture long-term dependencies in sequences, which is important for tasks such as speech recognition and natural language processing [4,5].

## Conclusion

LSTM networks have been successfully applied to various tasks such as speech recognition, natural language processing and time series prediction. In speech recognition, LSTM networks have been used to recognize phonemes and words from speech signals. In natural language processing, they have been used for tasks such as language modeling, sentiment analysis and machine translation. In time series prediction, they have been used to predict stock prices, weather patterns and other time-varying phenomena.

## Acknowledgement

None.

## Conflict of Interest

There are no conflicts of interest by author.

## References

1. Ma, Yinglong, Xiaofeng Liu, Lijiao Zhao and Yue Liang, et al. "Hybrid embedding-

- based text representation for hierarchical multi-label text classification." *Expert Syst Appl* 187 (2022): 115905.
2. Liu, Minqian, Lizhao Liu, Junyi Cao and Qing Du. "Co-attention network with label embedding for text classification." *Neurocomputing* 471 (2022): 61-69.
  3. Xiao, Yaoqiang, Yi Li, Jin Yuan and Songrui Guo, et al. "History-based attention in Seq2Seq model for multi-label text classification." *Knowl Based Syst* 224 (2021): 107094.
  4. Jang, Joel, Yoonjeon Kim, Kyoungcho Choi and Sungho Suh. "Sequential targeting: A continual learning approach for data imbalance in text classification." *Expert Syst Appl* 179 (2021): 115067.
  5. Liu, Huiting, Geng Chen, Peipei Li and Peng Zhao, et al. "Multi-label text classification via joint learning from label embedding and label correlation." *Neurocomputing* 460 (2021): 385-398.

**How to cite this article:** Crooks, James. "Long Short-Term Memory Networks: Overcoming Vanishing Gradient Problem in Recurrent Neural Networks." *Int J Sens Netw Data Commun* 12 (2023): 212.