

Chapter 4


A Deep Understanding of Long Short-Term Memory for Solving Vanishing Error Problem: LSTM-VGP

Aswathy Ravikumar

 <https://orcid.org/0000-0003-0897-6991>

Vellore Institute of Technology, Chennai, India

Harini Sriraman

 <https://orcid.org/0000-0002-2192-8153>

Vellore Institute of Technology, Chennai, India

ABSTRACT

Long Short-Term Memory (LSTM) is a specific kind of recurrent neural network (RNN) structure that addresses the constraints of conventional RNNs in effectively capturing and learning long-term relationships in sequential input. In this chapter, the authors examine the LSTM cell and its modifications to investigate the LSTM cell's capability for learning. Furthermore, future study prospects for LSTM networks are outlined. LSTM networks have gotten extensive attention in scientific papers, technical websites, and deployment manuals because of their efficacy in a variety of practical situations. Gradient-based learning techniques used in RNNs are too slow because as the error is transmitted back, it disappears, resulting in a much more extended learning period. LSTMs handle the issue with a novel additive gradient design that incorporates direct access towards the forget gate's activations, allowing the network to promote desirable behavior from the error gradient by updating the gates often at each time step of learning.

INTRODUCTION

Deep Neural Networks, a subfield of machine learning that uses associations learned from data to make judgments (Nweke et al., 2018). This is a promising area of machine learning that can produce outcomes

DOI: 10.4018/978-1-6684-8531-6.ch004

A Deep Understanding of Long Short-Term Memory

that are more and more agreeable as data is gathered (LeCun et al., 2015).

Among the most popular DL algorithms is CNN. LeCun et al. (1989) presented this model in 1989, and it performed well in computer vision. Businesses such as Google, Apple, and Instagram are among those conducting ongoing research and integrating CNN to their services, in addition to academic institutions (Khan et al., 2020). Convolutional layer, down - sampling layer (also known as a pooling layer), plus fully linked layer are the three layers that make up the CNN model. Regional receptive fields with weight sharing are applied by the convolution and sub-sampling levels. They can be layered in various levels, with the final stage of categorization being carried out by a completely linked layer. Regarding feature extraction and classification, CNN is fantastic. In machine vision, categorization, and video recognition, it has risen to the top (John et al., 2021; Ravikumar & Sriraman, 2023; Zhang et al., 2018).

An effective learning paradigm when handling sequential data, such as voice recognition and language understanding, is the RNN, one of the promising DL models. Through the inner representation of a neural network's recollection of prior inputs, it develops characteristics for time information. RNN can also forecast information based on historical and current data. However, due to the gradient disappearing or gradient exploding issue, it is challenging to learn stored data over an extended period of time using the RNN structure (Gelenbe, 1993; Robin et al., 2021).

The LSTM model, first presented in 1997 (Hochreiter & Schmidhuber, 1997), provides a solution to this RNN problem in its core. By utilizing numerous gate components, LSTM cells are able to extract information, remember the information, or delete information stored that is no longer be required. As previously indicated, this model has many traits with RNN while also addressing some of its shortcomings. It can be applied in areas where sequential analysis data and event forecasting utilizing current data are necessary. In other respects, LSTM is among the most sophisticated time-series handling networks.

Since the first LSTM study published in 1997 (Hochreiter & Schmidhuber, 1997), several theoretical and practical publications have been released on the topic of this sort of RNN, with many describing the remarkable results produced in a broad range of sequential data application fields. The LSTM has exerted a significant influence on language modeling and speech-to-text conversion. machine translation, and more uses. Many academic and business readers, encouraged by the remarkable benchmarks given in the literature, desire to understand LSTM network to assess its relevance to their own study or practical use-case. Numerous RNN and LSTM network designs are efficiently and production-ready implemented in all of the main open-source platforms for deep learning.

Obviously, some professionals, even if they are unfamiliar with RNN/LSTM methods, take full advantage of such an accessibility and cost-effectiveness and immediately begin research and experimenting. Others want a comprehensive understanding of the working of this beautiful and successful system. The benefit of this longer path in that it effectively gives the opportunity to develop a degree of intuition that can prove useful all during phases of the procedure of integrating an open-source component to meet the requirements of their research program or productivity applications, including data preparation, troubleshooting, and tuning.

RNN

The fact that RNNs are the sole neural net type including an internal storage makes them among the most viable algorithms currently in use. RNN is a sort of neural network in which the output of the preceding step is used as input for the current phase. In conventional neural networks, all inputs and outputs

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-deep-understanding-of-long-short-term-memory-for-solving-vanishing-error-problem/335184

Related Content

A Fuzzy-GA for Predicting Terrorist Networks in Social Media

Amit Kumar Mishra, Vikram Rajpoot, Ramakant Bhardwaj, Pankaj Kumar Mishra and Pushpendra Dwivedi (2022). *Dark Web Pattern Recognition and Crime Analysis Using Machine Intelligence* (pp. 126-160).

www.irma-international.org/chapter/a-fuzzy-ga-for-predicting-terrorist-networks-in-social-media/304207

DFC: A Performant Dagging Approach of Classification Based on Formal Concept

Nida Meddouri, Hela Khoufi and Mondher Maddouri (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 38-62).

www.irma-international.org/article/dfc/277433

Harnessing Environmental Intelligence to Enhance Crop Management by Leveraging Deep Learning Technique

S. Gandhimathi Alias Usha (2023). *Machine Learning and Deep Learning for Smart Agriculture and Applications* (pp. 106-123).

www.irma-international.org/chapter/harnessing-environmental-intelligence-to-enhance-crop-management-by-leveraging-deep-learning-technique/329892

Sentiment Analysis of Game Review Using Machine Learning in a Hadoop Ecosystem

Arvind Panwar and Vishal Bhatnagar (2020). *Handbook of Research on Engineering Innovations and Technology Management in Organizations* (pp. 145-165).

www.irma-international.org/chapter/sentiment-analysis-of-game-review-using-machine-learning-in-a-hadoop-ecosystem/256674

Self-Driving Cars: Role of Machine Learning

Pooja Jha and K. Sridhar Patnaik (2020). *Handbook of Research on Emerging Trends and Applications of Machine Learning* (pp. 490-507).

www.irma-international.org/chapter/self-driving-cars/247578