

Long Short-Term Memory Network (LSTM)

Data Mining & Neural Networks

Dr. Wilmer Garzón

Director, Master's Program in Data Science
Department of Computer Engineering

Escuela Colombiana de Ingeniería
Universidade da Coruña

2025



UNIVERSIDAD



Neural Networks LSTM

Introduction to the topic of **Long Short-Term Memory (LSTM) networks** and their significance in modeling sequential data.

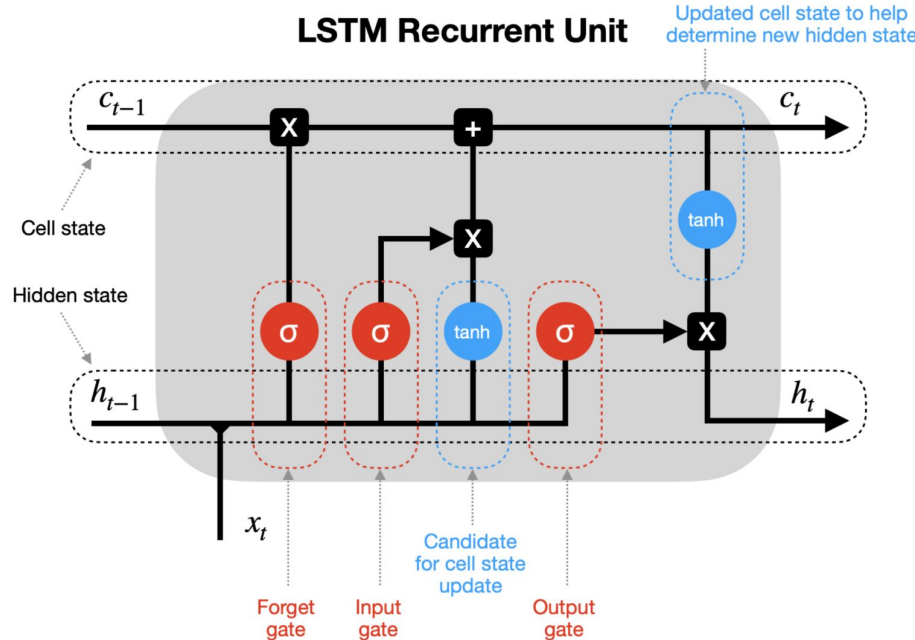


Motivation for LSTMs

- Traditional RNNs struggle with learning **long-term dependencies** due to vanishing or exploding gradients during backpropagation.
- LSTMs were introduced to address this limitation by **incorporating a memory cell** that can maintain its state over time.

LSTM Architecture Overview

An LSTM unit consists of a **cell state** and three gates: **input gate**, **forget gate**, and **output gate**.



LSTM Architecture Components



- The **cell state** is the core component of an LSTM.
- It acts like a **conveyor belt**, carrying information across time steps with minimal changes.
- This allows the network to **maintain long-term memory**:
 - **Forget gate** decides what information to **discard**.
 - **Input gate adds new** relevant **information**.
 - **Output gate** controls what **part of the cell state**.
- **LSTMs retain important information over long sequences.**



LSTM vs. Traditional RNNs

- Traditional RNNs suffer from short-term memory and vanishing gradients.
- LSTMs are designed to remember information over long sequences.

Applications of LSTMs

- **Natural Language Processing:** text generation, sentiment analysis, machine translation
- **Speech Recognition:** converting audio to text
- **Time Series Forecasting:** stock prices, weather prediction
- **Anomaly Detection:** identifying unusual patterns in data
- **Music Generation:** composing melodies based on learned patterns



Some LSTM Approaches:

- **NLP: LSTMs are used to model sequences of words or characters.**
 - They can **capture context and dependencies** across long distances.
- **LSTM in Time Series:**
 - LSTMs are well-suited for time series forecasting due to their ability to model temporal dependencies.
- **LSTM in Anomaly Detection:**
 - LSTMs can be used to **detect anomalies in sequential data** by learning normal patterns.

Challenges with LSTMs

- **Training Complexity:** LSTMs require more time and resources to train compared to simpler models.
- **Sequential Processing:** Their architecture limits parallelization, making them slower on large datasets.
- **Overfitting Risk:** With many parameters, LSTMs can easily overfit small datasets.
- **Difficult to Interpret:** The internal workings of gates and memory states are not easily explainable.

Comparing CNN, RNN, and LSTM

<u>Feature</u>	<u>CNN</u>	<u>RNN</u>	<u>LSTM</u>
Best for	Spatial data (e.g., images)	Sequential data (e.g., text)	Long-term sequential data
Memory	No memory	Short-term memory	Long-term memory via cell state
Handles sequences	No	Yes	Yes (better than RNN)
Gradient issues	Not applicable	Suffers from vanishing gradients	Designed to avoid this problem
Parallelization	Highly parallelizable	Sequential processing	Sequential, slower than CNN
Use cases	Image classification, object detection	Language modeling, time series	Translation, speech recognition

Comparing LSTMs vs LLMs

LLMs offer powerful language capabilities but require significant data and compute.

Feature	LSTM (Long Short-Term Memory)	LLM (Large Language Model)
Architecture	Recurrent Neural Network (RNN)	Transformer-based
Memory Mechanism	Cell state with gates (input, forget, output)	Self-attention over large context windows
Training Data	Smaller datasets, sequential input	Massive datasets (text, code, etc.)
Parallelization	Limited (sequential processing)	High (parallelizable attention mechanism)
Model Size	Small to medium	Very large (billions of parameters)
Use Cases	Time series, speech recognition, basic NLP	Chatbots, text generation, translation, summarization
Strength	Captures temporal dependencies effectively	Deep contextual understanding and generalization

Summary and Conclusion

- LSTMs are powerful **tools for modeling sequential data and capturing long-term dependencies**.
- Their architecture **with memory cells and gating mechanisms enables** them to outperform traditional RNNs in many applications.

LLMs overcome LSTM limitations with broader context and parallel processing



UNIVERSIDAD