



LSTM (Long Short-Term Memory)

Agenda

KEY TOPICS DISCUSSED IN THIS PRESENTATION

- Understanding LSTM
- Sequential Data
- Recurrent Neural Network (RNN) in LSTM
- How LSTM works and its structure
- Forget gate, Input gate and Output gate
- Application of LSTM
- Optimization and Limitation of LSTM

Understanding LSTM

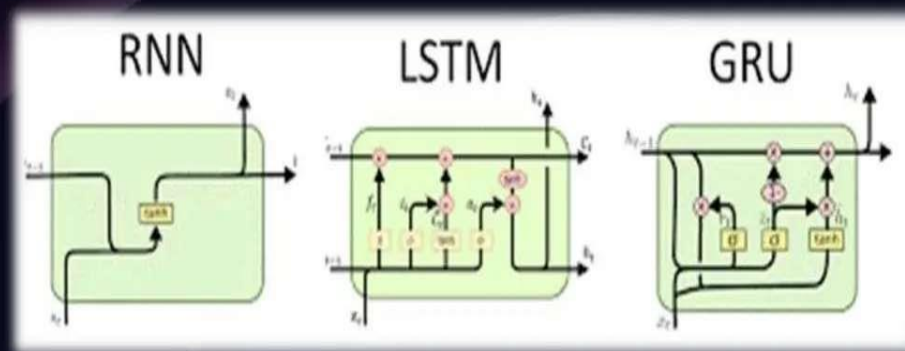
- Unlocking the Power of Long Short-Term Memory Networks

A Long Short Term Memory Network (LSTM) is a type of recurrent neural network that can learn long-term dependencies and is widely used in applications such as Natural Language Processing, Speech Recognition, and Time Series Prediction.

Sequential Data

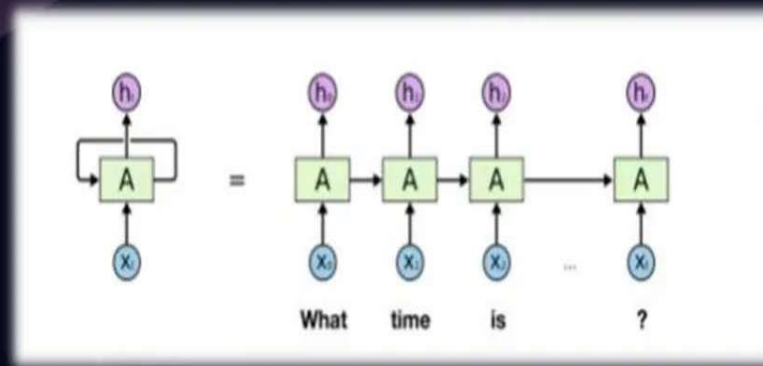
- An LSTM network enables us to input sequence data into a network, and make predictions based on the individual time steps of the sequence data.
- Sequential data includes text streams, audio clips, video clips, time-series data and etc. Recurrent Neural Networks (RNNs) is a popular algorithm used in sequence models.

How to Handle Sequential Data



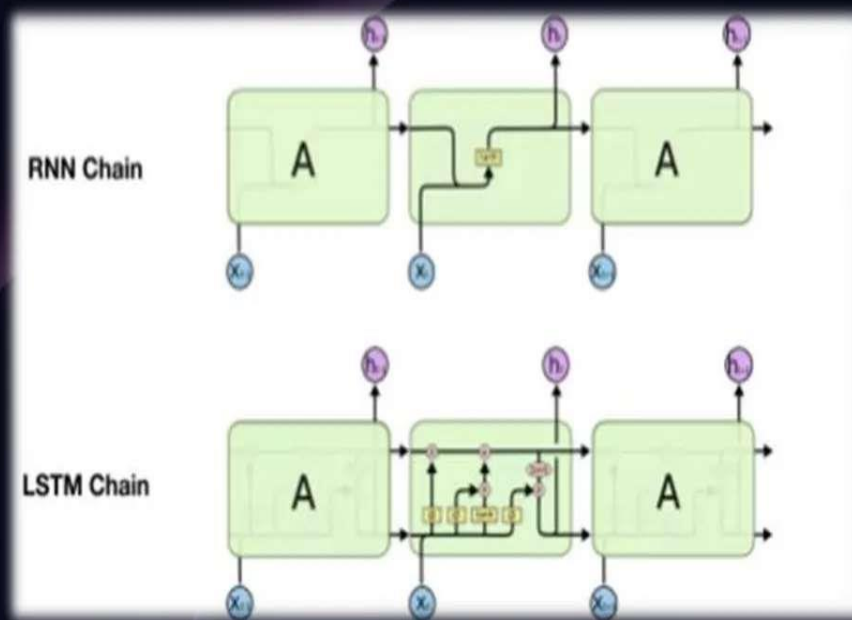
- Text, Stock prices, Sensor signals, DNA, Customer purchase behavior, Sound signals.
- Bag of words doesn't preserve order/sequence in data
- Modeling sequential data requires a temporal architecture in data

RNN in LSTM



- LSTM is a type of RNN with higher memory power to remember the outputs of each node for a more extended period to produce the outcome for the next node efficiently.
- RNN has a memory that stores all information about the

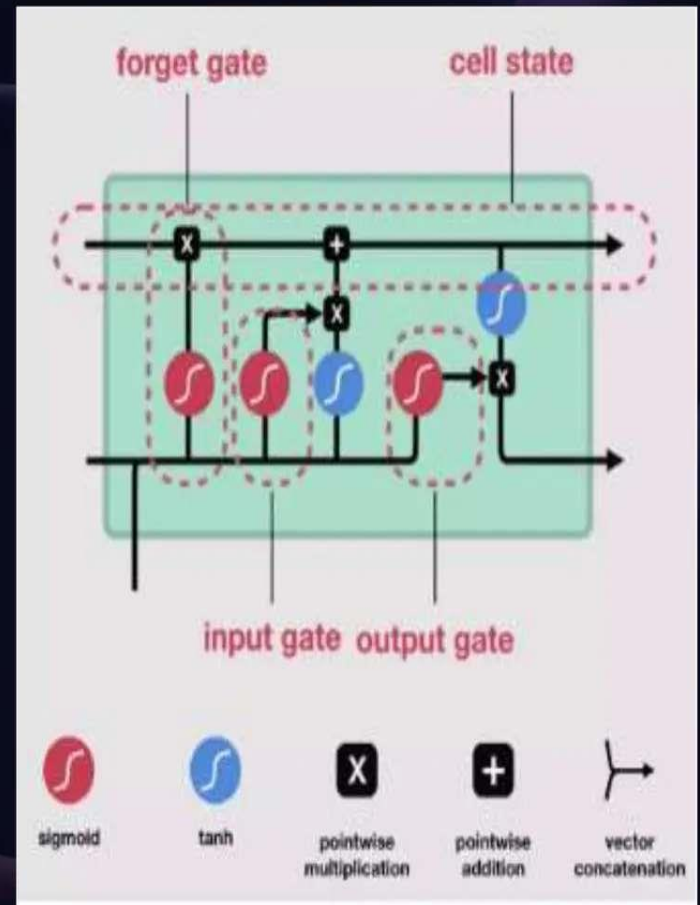
How do LSTM Work



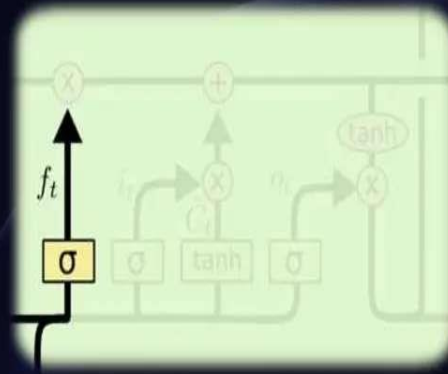
- LSTM try to add long term memory to remember certain hidden states more than others. This allows them to retain knowledge sequences.
- They have 2 outputs instead of 1, the hidden and the cell state. Their

Structure of LSTM

- An LSTMs architecture consists of 3 gates – Forget gate, Input Gate, Output Gate
- Tanh acts as a squashing function while sigmoid acts as a decision function (gate)
- Cell state is a channel that runs along the LSTM chain carrying information from one time-step to another freely



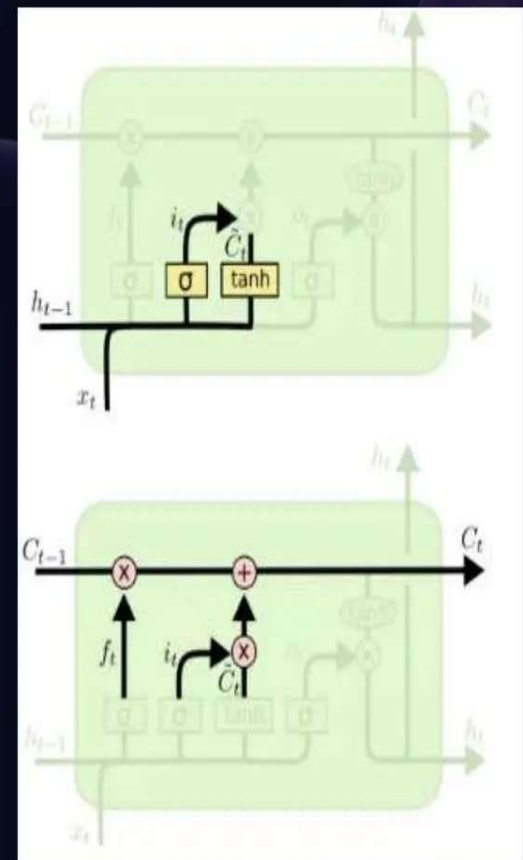
Forget Gate



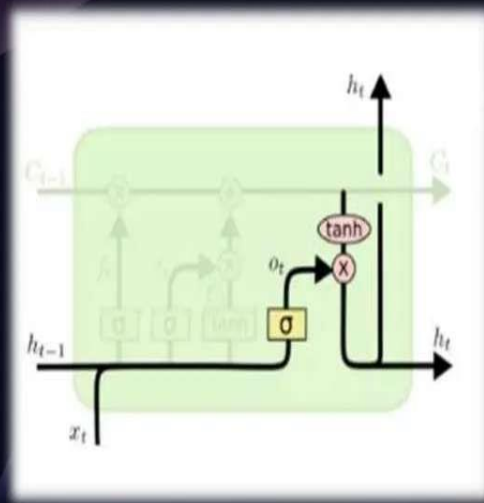
- It concatenates the previous hidden state to the current input, multiplies it with weights and adds a bias, then applies a sigmoid function before multiplying it to the cell state.

Input Gate

- Input gate decides what information needs to be saved to the cell state, it simply does the same operation as the forget gate but instead of writing on to the cell state it combines (multiplies) it with the Tanh of the concatenated vector of hidden state and input. This is then added to the cell state, which has been updated by the forget gate already



Output Gate



- The output gate determines the value of the next hidden state. This state contains information on previous inputs. First, the values of the current state and previous hidden state are passed into the third sigmoid function. Then the new cell state generated from the cell state is passed through the tanh function

Applications of LSTM

Natural Language Processing

LSTM models are widely used for language modeling, text classification, machine translation, and sentiment analysis.

Speech Recognition

LSTM can be trained to recognize and transcribe speech, enabling applications like voice assistants and transcription services.

Time Series Prediction

LSTM can accurately forecast future values in time series data, making it valuable for financial predictions, weather forecasting, and more.

Training of LSTM

- Backpropagation Through Time (BPTT): The algorithm used to compute gradients in LSTM networks.
- Train with time series data to remember and predict based on long term dependencies.

Limitations and Challenges

- Vanishing Gradient Problem: LSTM networks can suffer from the vanishing gradient problem, making it difficult to learn long-term dependencies.
- Overfitting: LSTM models are prone to overfitting when the training data is limited or noisy.

Future Directions and Advancements

Researchers are actively working on improving LSTM models by developing variants such as Gated Recurrent Units (GRUs) and exploring techniques like attention mechanisms and self-attention for better performance. The future looks promising for LSTM and its applications in various domains.

Conclusion

Long Short Term Memory Networks are powerful tools in the field of machine learning, enabling the modeling of complex temporal dependencies. Their applications in Natural Language Processing, Speech Recognition, and Time Series Prediction continue to drive advancements in the field, with ongoing research focused on addressing their limitations and improving performance.



Thank You!

Q&A?