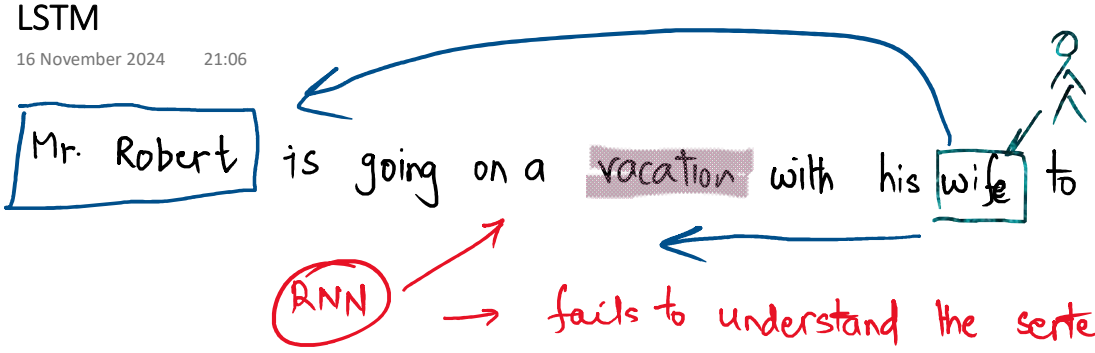
LSTM

16 November 2024 21:06



→ fails to understand the sentence fully

↓

RNN has short-term memory

GHAJINI → 15 mins)

* [Reason → Vanishing gradients]

Chain rule in differentiation

slope at each step

$(0.2) \times (0.1) \times (0.3) \times$ ------ 9 times

$= (0.1)^9 \rightarrow 0.000000001 = 1 \times 10^{-9} \cong 0 \rightarrow$ 'gradient is vanishing'

During backpropagation, gradients of the loss function become very small as they are propagated through time, especially for long sentences

This results in very slow or negligible updates to earlier layers in the network.

Effect: RNN struggles to learn dependencies over long sequences.

Exploding gradient problem

Gradients can grow exponentially large during back propagation, leading to numerical instability.

Effect: Training becomes unstable, and the network fails to converge.

Long-term dependencies

Standard RNNs have a limited ability to remember information from earlier time steps due to the issues highlighted above.
RNN has difficulty with long-term dependencies.

Effect: RNNs tend to forget important information from the distant past in a sequence.

LSTM

Long short-Term Memory is a special type of RNN architecture designed to handle sequential data and long-term dependencies effectively.

LSTMs were introduced by Hochreiter and Schmidhuber in 1997 to address the limitations of std. RNNs particular by the vanishing gradient problem

Key features of LSTM

1. Memory cells

The central component of an LSTM is the memory cell which retains information across long sequences

And, the memory cell decides what information (1) to keep, (2) update, or (3) discard using gating mechanisms.

(1) KEEP → (2) UPDATE → (3) DISCARD.

2. Gates

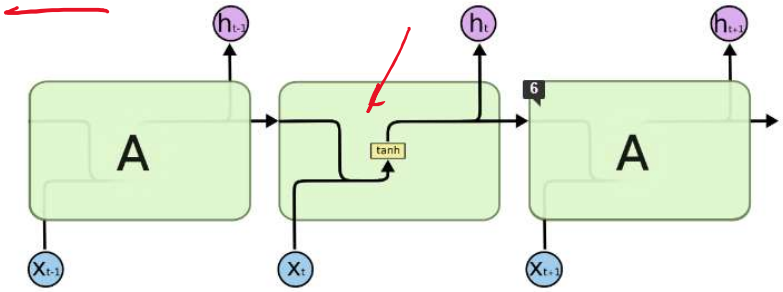
LSTM's use gates to control the flow of information - These are:

→ Forget Gate: decides what information to discard from the cell state

→ Input Gate: decides what new information to add to the cell state

→ Output Gate: decides what part of the cell state to output as the hidden state.

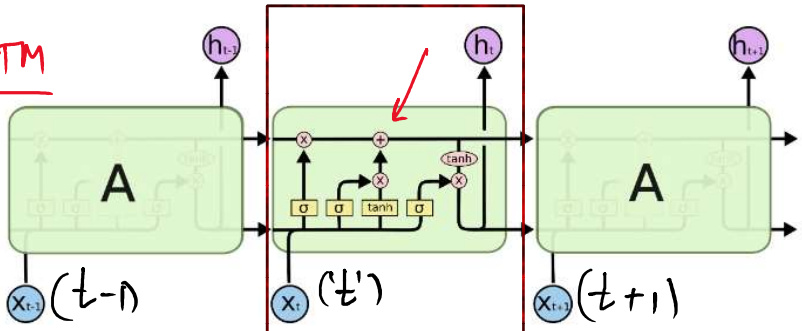
RNN



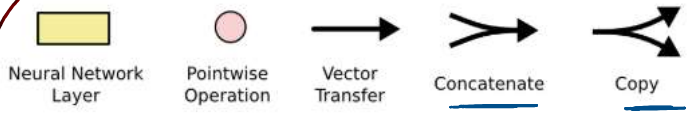
The repeating module in a standard RNN contains a single layer.

Vs

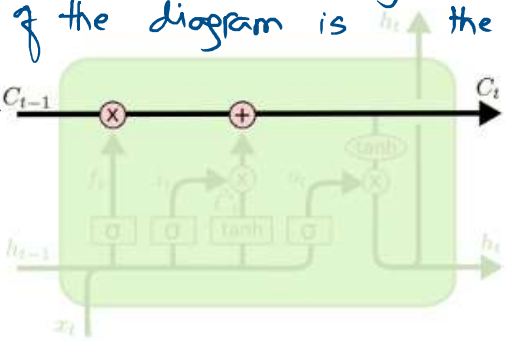
LSTM



The repeating module in an LSTM contains four interacting layers.

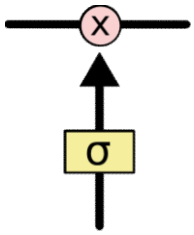


horizontal line running through top of the diagram is the cell state → is like a conveyor belt



LSTM does have the ability to remove or add information to the cell state, carefully regulated by structure called gates.

Gates



they are composed out of a sigmoid neural net layer and a pointwise multiplication operation

sigmoid layer outputs between 0 – 1 describing

sigmoid layer outputs between 0 – 1 describing how much of each component should be let through.

sigmoid value = 0 → do not let anything through

sigmoid value = 1 ⇒ let everything through.