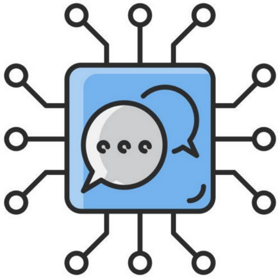
GRU and LSTM

Tushar B. Kute,
http://tusharkute.com



Problems with RNN

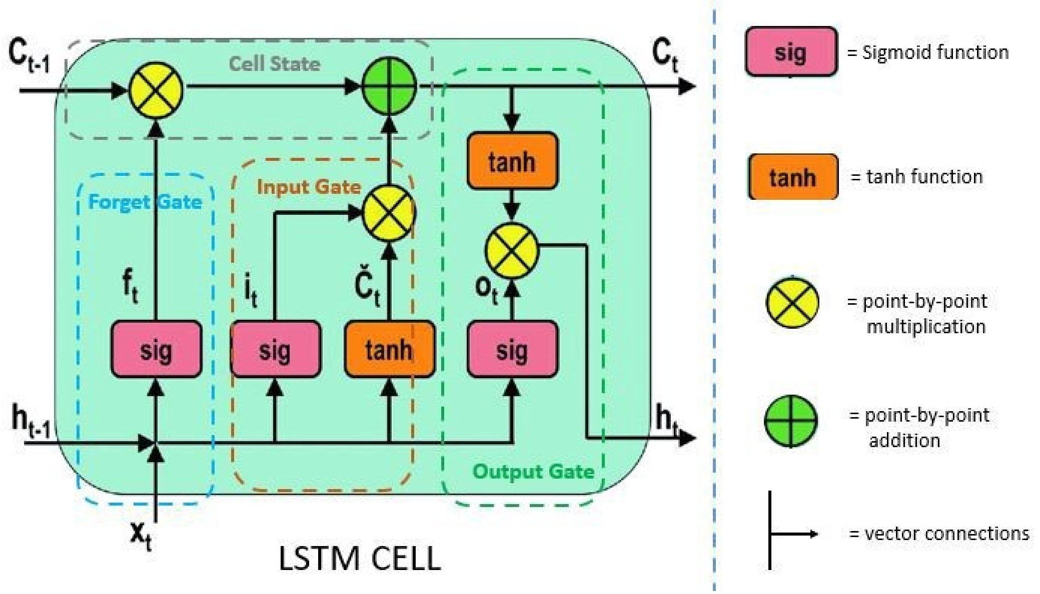
- Exploding and vanishing gradient problems during backpropagation.
- Gradients are those values which to update neural networks weights. In other words, we can say that Gradient carries information.
- Vanishing gradient is a big problem in deep neural networks. it vanishes or explodes quickly in earlier layers and this makes RNN unable to hold information of longer sequence. and thus RNN becomes short-term memory.
- If we apply RNN for a paragraph RNN may leave out necessary information due to gradient problems and not be able to carry information from the initial time step to later time steps.

How to solve these problems?

- The reason for exploding gradient was the capturing of relevant and irrelevant information. a model which can decide what information from a paragraph and relevant and remember only relevant information and throw all the irrelevant information
- This is achieved by using gates. the LSTM (Long -short-term memory) and GRU (Gated Recurrent Unit) have gates as an internal mechanism, which control what information to keep and what information to throw out. By doing this LSTM, GRU networks solve the exploding and vanishing gradient problem.
- Almost each and every SOTA (state of the art) model based on RNN follows LSTM or GRU networks for prediction.

LSTM Networks

- Every LSTM network basically contains three gates to control the flow of information and cells to hold information. The Cell States carries the information from initial to later time steps without getting vanished.



LSTM Networks

- Forget Gate:
  - This gate decides what information should be carried out forward or what information should be ignored.
  - Information from previous hidden states and the current state information passes through the sigmoid function. Values that come out from sigmoid are always between 0 and 1.
  - if the value is closer to 1 means information should proceed forward and if value closer to 0 means information should be ignored.

LSTM Networks

- Input Gate:
  - After deciding the relevant information, the information goes to the input gate, Input gate passes the relevant information, and this leads to updating the cell states. simply saving updating the weight.
  - Input gate adds the new relevant information to the existing information by updating cell states.

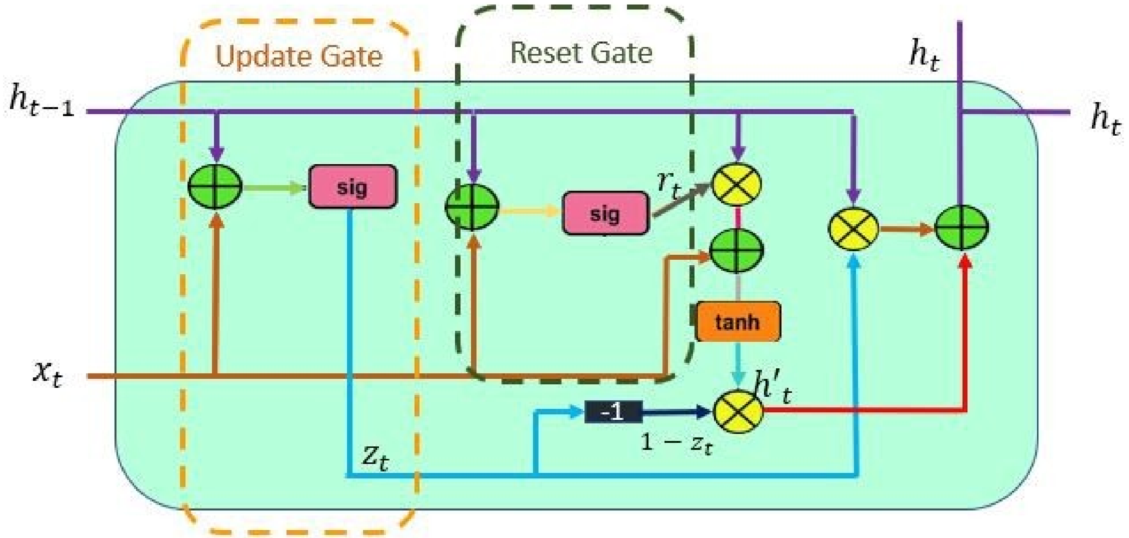
LSTM Networks

- Output Gate:
  - After the information is passed through the input gate, now the output gate comes into play.
  - Output gate generates the next hidden states. and cell states are carried over the next time step.

GRU

- GRU (Gated Recurrent Units) are similar to the LSTM networks. GRU is a kind of newer version of RNN. However, there are some differences between GRU and LSTM.
  - GRU doesn’t contain a cell state
  - GRU uses its hidden states to transport information
  - It Contains only 2 gates(Reset and Update Gate)
  - GRU is faster than LSTM
  - GRU has lesser tensor’s operation that makes it faster

GRU



Gated Recurrent Network (GRU)

sig = Sigmoid function

tanh = tanh function

= Hadamard Product operation

= addition operation

= vector connections

GRU

- 1. Update Gate
  - Update Gate is a combination of Forget Gate and Input Gate. Forget gate decides what information to ignore and what information to add in memory.
- 2. Reset Gate
  - This Gate Resets the past information in order to get rid of gradient explosion. Reset Gate determines how much past information should be forgotten.

Examples:

- Practical

Thank you

@mitu_skillologies

@mITuSkillologies

@mitu_group

@mitu-skillologies

@MITUSkillologies

kaggle

@mituskillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

@mituskillologies

contact@mitu.co.in

tushar@tusharkute.com