Summary of the Paper

"Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling"

Link:- [1412.3555]

Problem:-

Standard RNNs struggle to learn long-term dependencies due to vanishing gradients. Though LSTMs address this with complex gating mechanisms, their efficiency and necessity remain underexplored.

Solution:-

The authors evaluate a simpler alternative, the Gated Recurrent Unit (GRU), comparing it empirically to both LSTMs and traditional RNNs on sequence modeling tasks (music and speech data).

Results:

GRUs perform comparably to LSTMs and significantly better than standard RNNs, while requiring fewer parameters and less computation.

Benefits:-

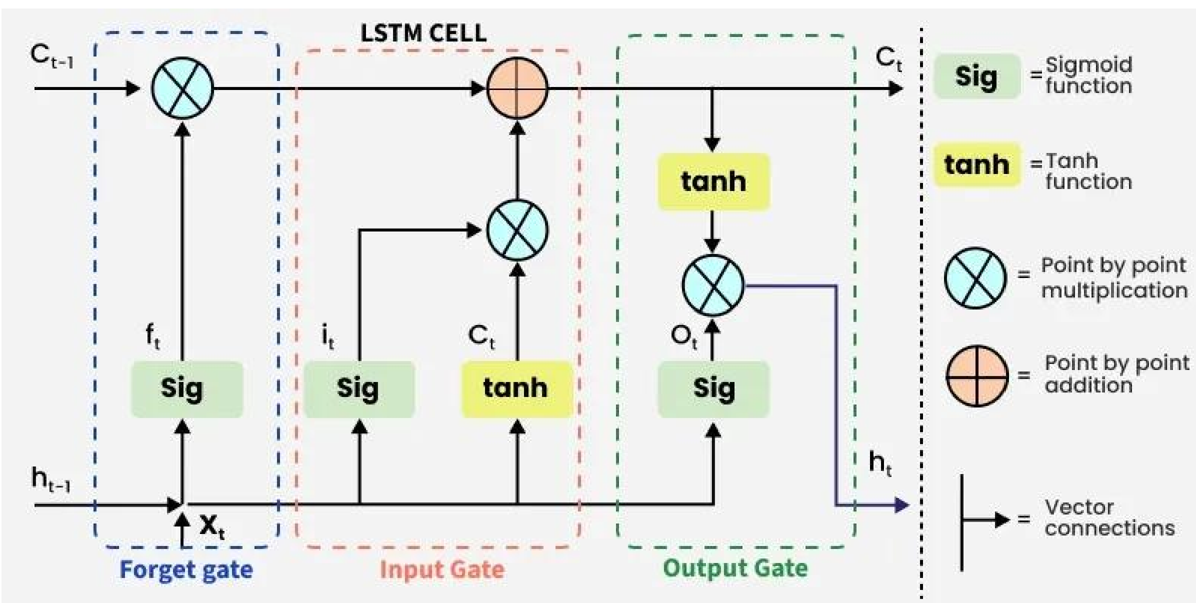
- GRU is simpler and faster than LSTM.
- Both GRU and LSTM handle long-term dependencies well.
- GRU is a strong candidate for efficient sequence modeling, especially in constrained environments.

LSTM

LSTM architecture involves the memory cell which is controlled by three gates: input gate, the forget gate and the output gate. These gates decide what information to add to, remove from and output from the memory cell.

- **Input gate:** Controls what information is added to the memory cell.
- **Forget gate:** Determines what information is removed from the memory cell.
- **Output gate:** Controls what information is output from the memory cell.

This allows LSTM networks to selectively retain or discard information as it flows through the network which allows them to learn long-term dependencies. The network has a hidden state which is like its short-term memory. This memory is updated using the current input, the previous hidden state, and the current state of the memory cell.



Disadvantage:- LSTMs are more complex than basic RNNs due to their multiple gates (input, forget, output), which increases training time and memory usage.

GRU

GRU is a type of recurrent neural network that simplifies the architecture of LSTM by combining the forget and input gates into a single update gate. It is designed to capture long-term dependencies while being computationally more efficient than LSTM.

It contains two gates:-

- **Update gate**:- This gate decides how much information from previous hidden state should be retained for the next time step.
- **Reset gate**:- This gate determines how much of the past hidden state should be forgotten.