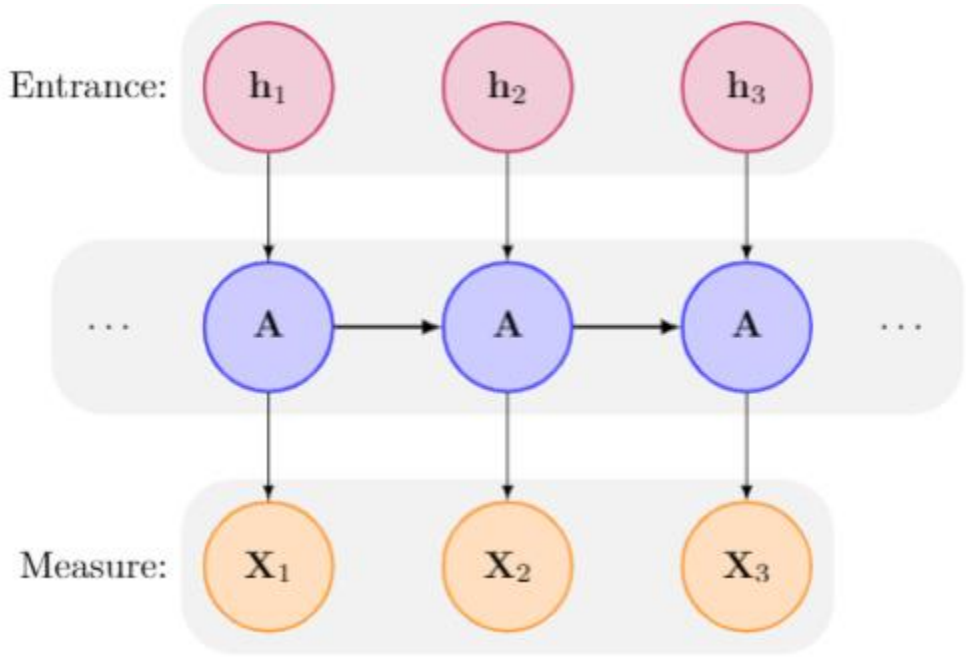
1.

(1)



(2)

The weight vector will be the same for all input which provide a O(1) space complexity as its equivalent to feed-forward neural network with O(t) space complexity.

2.

(1)

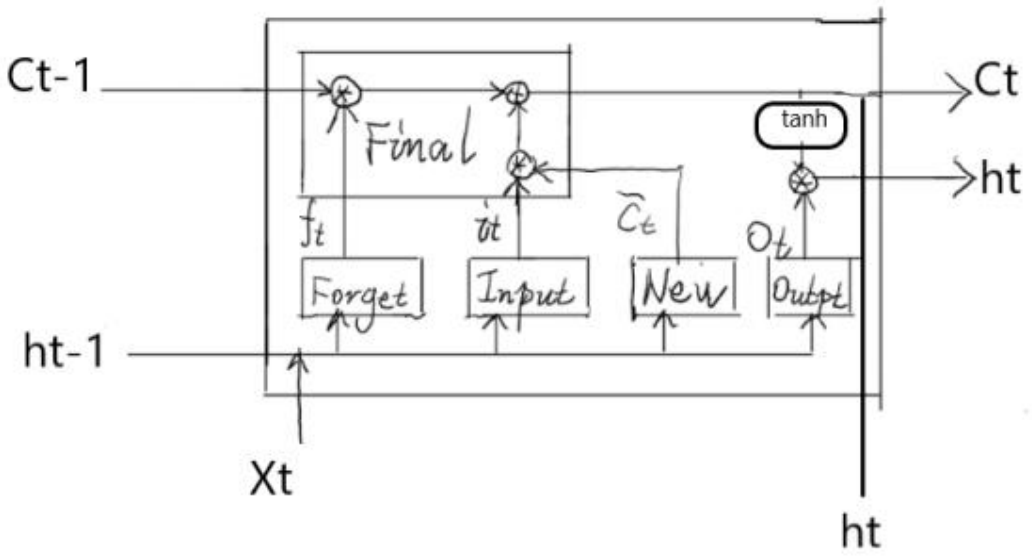
From 0 to 1

(2)

$\sum_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}} \leq (a_a a_b)^{t-k}$

- If $(a_a a_b) > 1$, then $\sum_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}}$ will goes ∞ and cause gradient explosion.
- If $(a_a a_b) < 1$, then $\sum_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}}$ will goes 0 and cause gradient vanishing.

3.



4.

The Forget Gate will not keep the cell state for those cells that's no need to be remembered and output a number close to 1 for those cell state that need to be memorized.

5.

LSTM RNNs replaces those neurons in the hidden layers with memory blocks (which contain memory cells that have recurrent connections).

Each cell has its own inputs, outputs and memory.

Cells that belong to the same block, share input, output and forget gates. This means that each cell might hold a different value in its memory, but the memory within the block is written to, read from and erased all at once.