|  |  |
| --- | --- |
| http://sphotos.ak.fbcdn.net/photos-ak-snc1/v2219/119/28/51288393663/n51288393663_1327303_6197.jpg | old_logo_color |

Report on Week8

By

Ahmed Usama Khalifa

Submitted to

Dr. Omar Nasr

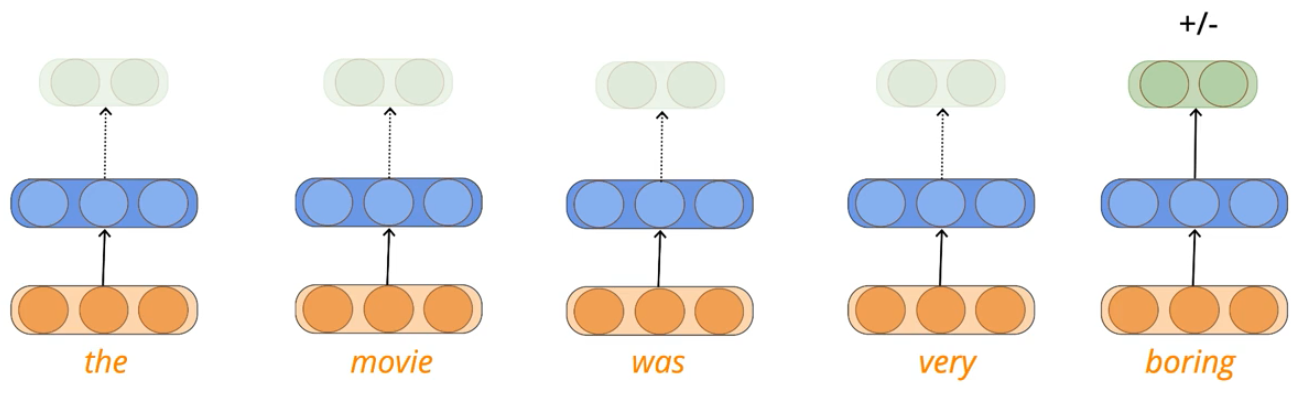
FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

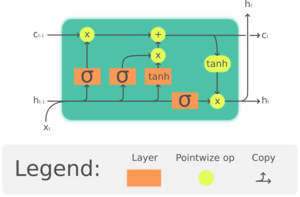
March – 2020

Sequence Models

**Sequence Modeling** is the task of predicting what word/letter comes next. Unlike the FNN and CNN, in **sequence modeling**, the current output is dependent on the previous input and the length of the input is not fixed.



LSTM

Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points, but also entire sequences of data.  


The Long Short-Term Memory (LSTM) cell can process data sequentially and keep its hidden state through time.

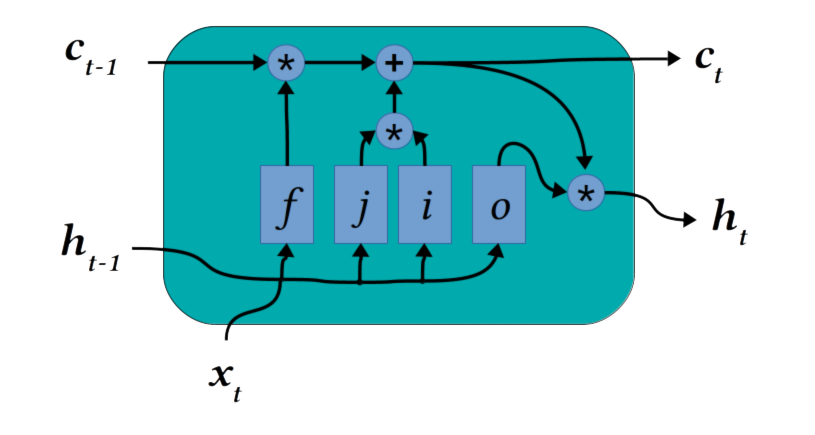
Types of LSTM

1. LSTM Classic

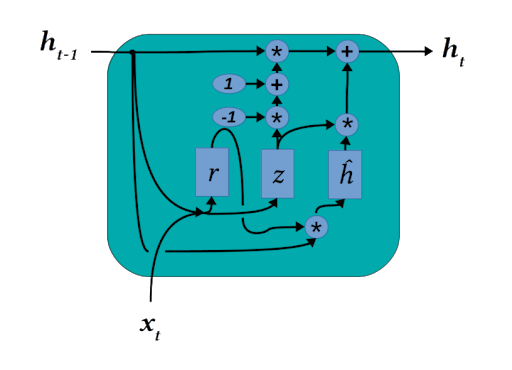
The classic LSTM architecture is characterized by a persistent linear cell state surrounded by non-linear layers feeding input and parsing output from it. Concretely the cell state works in concert with 4 gating layers, these are often called the forget, (2x) input, and output gates.

The forget gate chooses what values of the old cell state to get rid of, based on the current input data. The two input gates (often denoted *i* and *j*) work together to decide what to add to the cell state depending on the input. *i* and *j* typically have different activation functions, which we intuitively expect to be used to suggest a scaling vector and candidate values to add to the cell state.

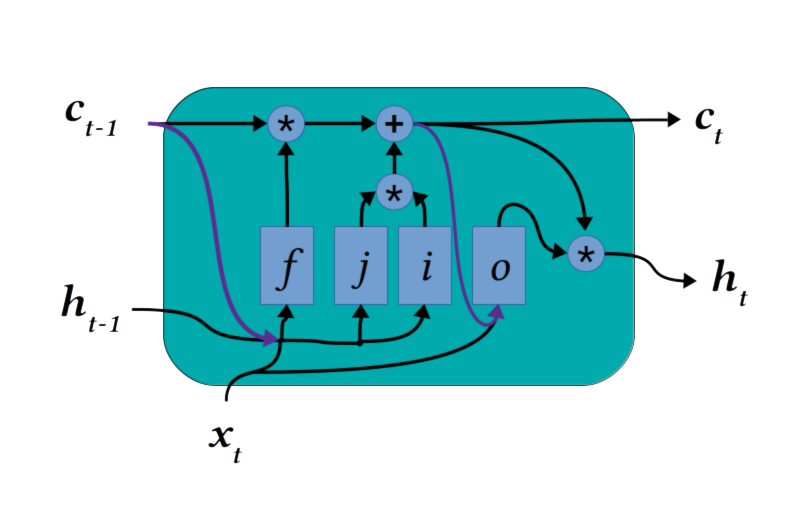
Finally, the output gate determines what parts of the cell state should be passed on to the output. Note that in the case of classic LSTMs, the output *h* consists of hidden layer activations (these can be subjected to further layers for classification, for example) and the input consists of the previous hidden state output and any new data *x* provided at the current time step.



## **2. Gated Recurrent Unit**



## **3. Peephole Connections**

4. Multiplication LSTM

5. LSTMs With Attention

Sequential modes

