

Introduction to NLP (CS7.401)

Spring 2022, IIIT Hyderabad
11 Mar, Friday (Lecture 13)

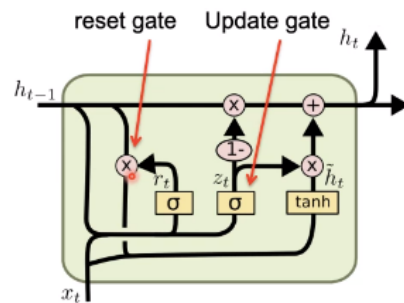
Taught by Prof. Manish Shrivastava

Neural Network Language Models

Long Short-Term Memory Networks

A modified version of LSTMs, the GRU (gated recurrent unit), merges the cell state c_t and hidden state h_t . It also combines the *forget* and *input* gates into a single *update* gate.

GRUs ensure that h_t either retains a high amount of old information or starts over with a high amount of new information.



$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{aligned}$$

Figure 1: Gated Recurrent Unit

Applications

Sequence to Sequence Chat Model

Neural Machine Translation

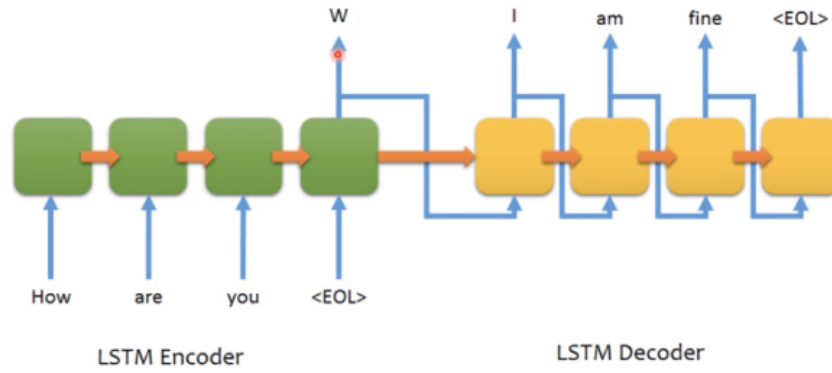


Figure 2: Sequence to Sequence Chat Model using LSTMs

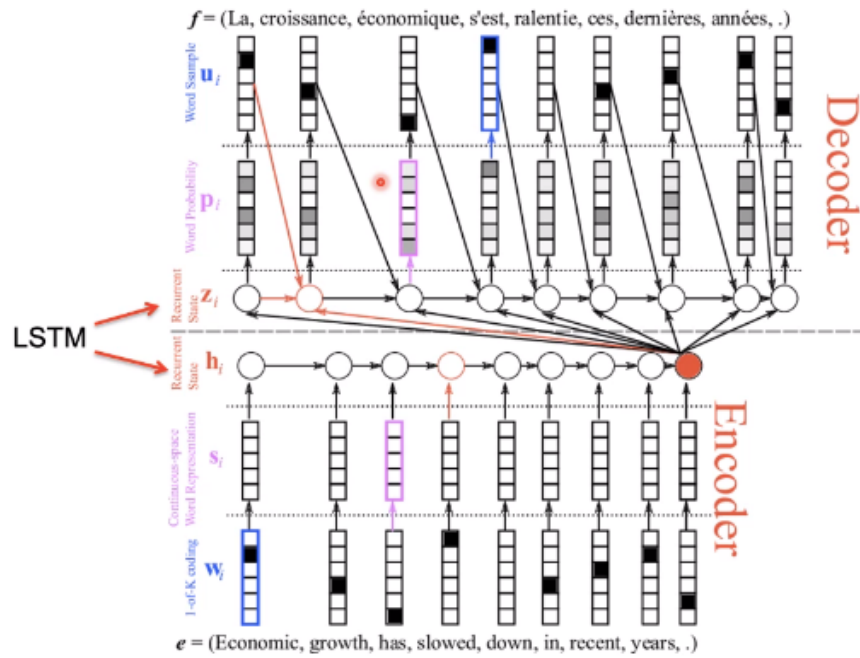


Figure 3: Neural Machine Translation using LSTMs