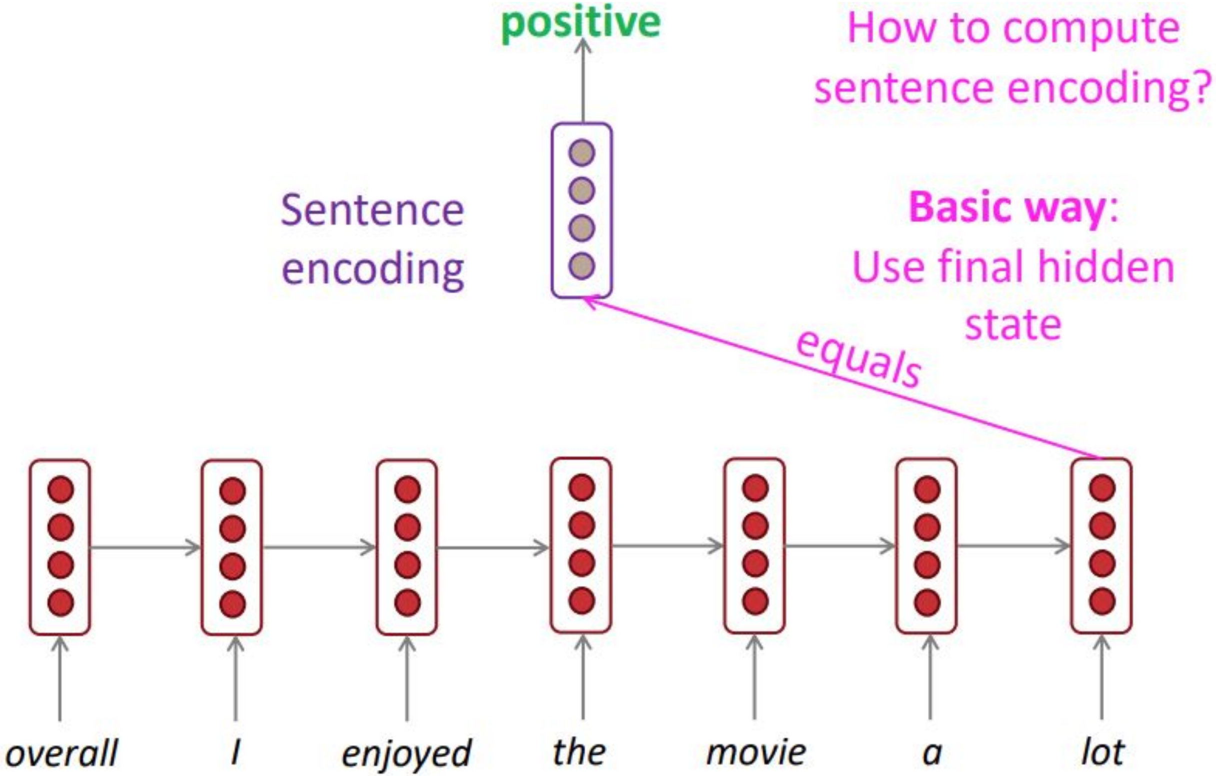
NLP and Deep Learning MAT3399

Lecture 6: Bi-directional LSTM & Sequence to Sequence

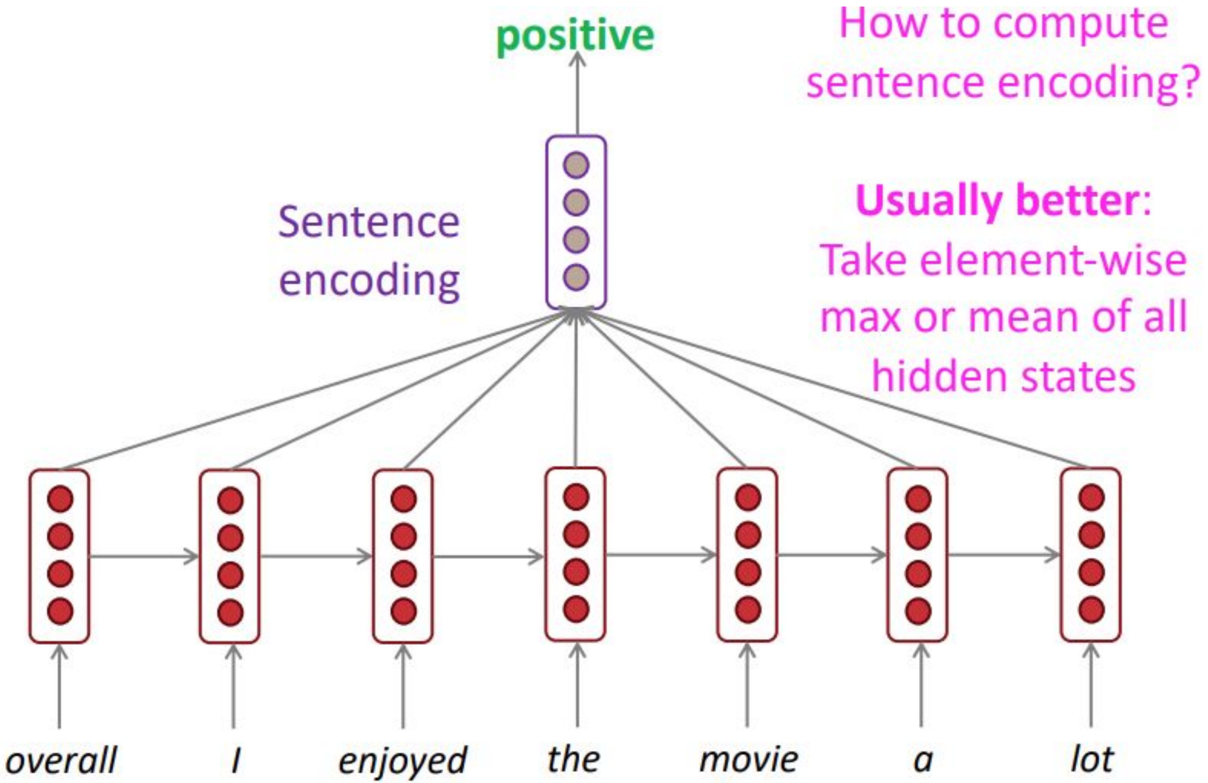
Tuan Anh Nguyen @ Aimesoft
ted.nguyen95@gmail.com

Content taken from CS224N Stanford

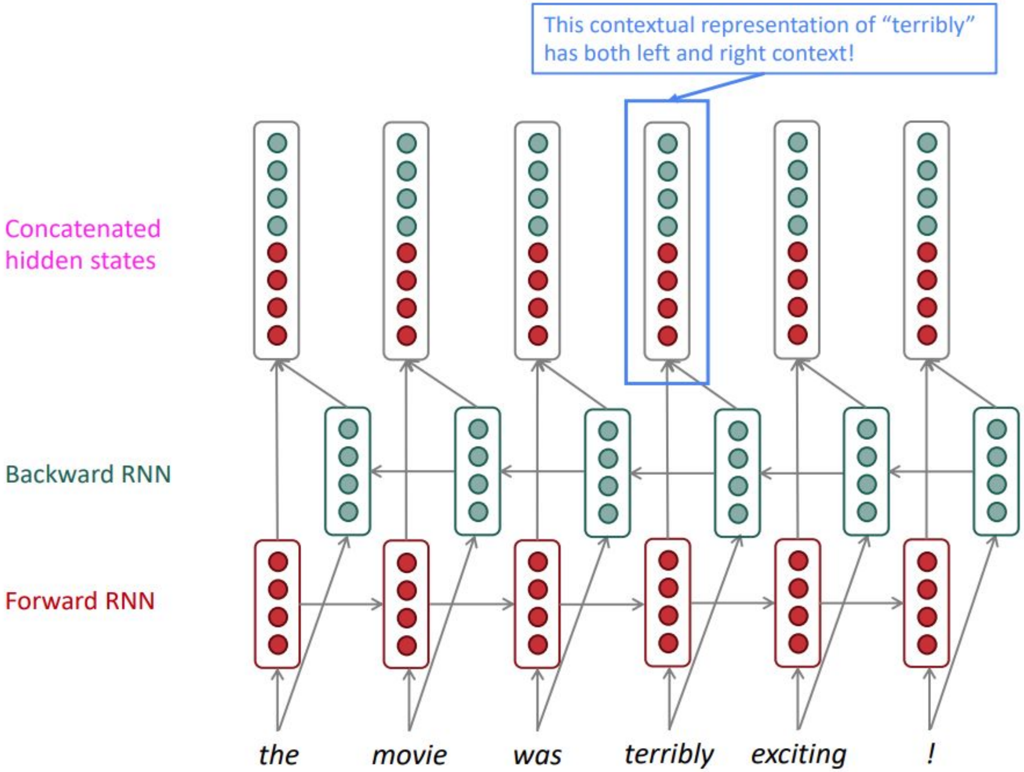
RNN can be used for sentence encoding



RNN can be used for sentence encoding



Bi-RNN / Bi-LSTM



Bidirectional RNNs are only applicable if you have access to the entire input sequence

Machine Translation

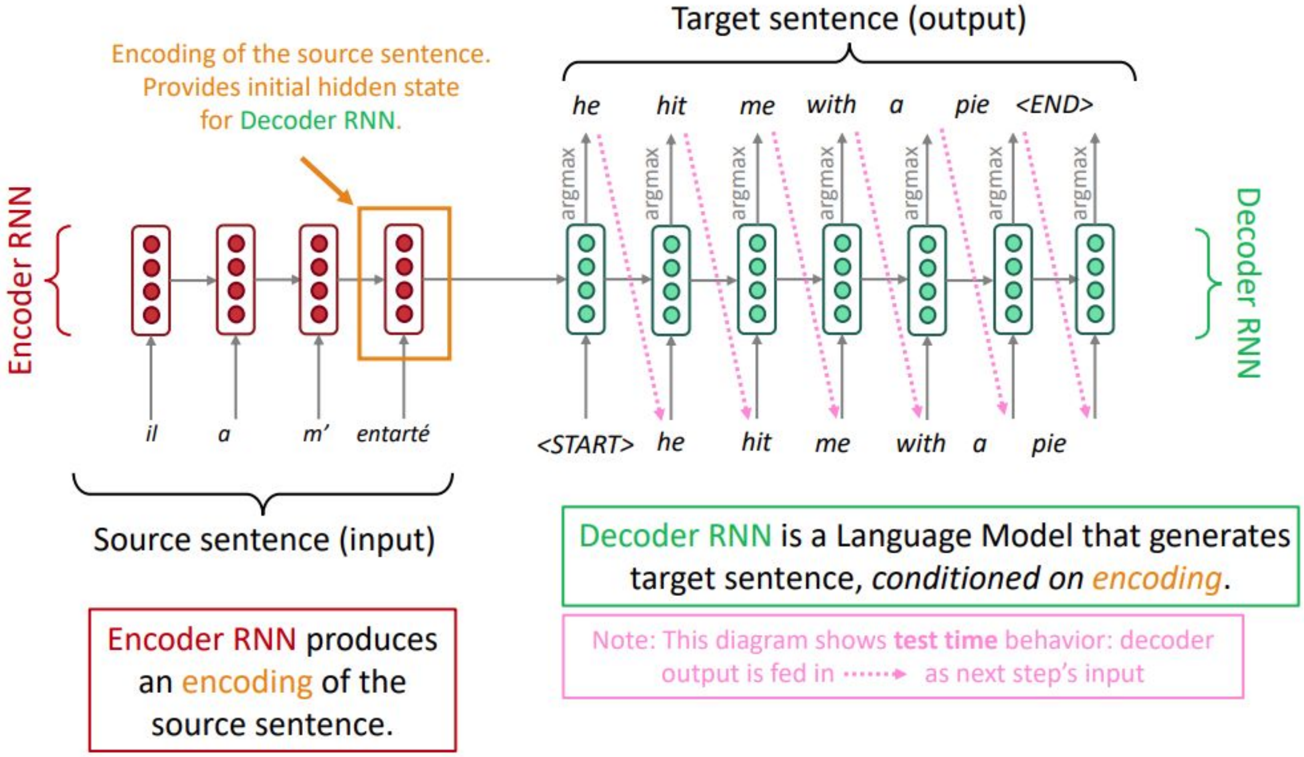
Machine Translation (MT) is the task of translating a sentence x from one language (the source language) to a sentence y in another language (the target language).

x: *L'homme est né libre, et partout il est dans les fers*

↓

y: *Man is born free, but everywhere he is in chains*

Sequence to Sequence Model for Machine Translation



Seq2seq model involves two RNNs:

- Encoder
- Decoder

Seq2seq can be used for many other tasks (E.g: Summarization, Dialogue System,...)

Evaluate Machine Translation Model – BLEU Score

$$\text{BLEU} = \underbrace{\min\Big(1, \exp\big(1 - \frac{\text{reference-length}}{\text{output-length}}\big)\Big)}_{\text{brevity penalty}} \underbrace{\Big(\prod_{i=1}^{4} precision_i\Big)^{1/4}}_{\text{n-gram overlap}}$$

with

$$precision_i = \frac{\sum_{\text{snt}\in\text{Cand-Corpus}} \sum_{i\in\text{snt}} \min(m^i_{cand}, m^i_{ref})}{w^i_t = \sum_{\text{snt'}\in\text{Cand-Corpus}} \sum_{i'\in\text{snt'}} m^{i'}_{cand}}$$

where

- m^i_{cand} is the count of i-gram in candidate matching the reference translation
- m^i_{ref} is the count of i-gram in the reference translation
- w^i_t is the total number of i-grams in candidate translation

Calculating BLEU Score – Example

Reference: `the cat is on the mat`
Candidate: `the the the cat mat`

The first step is to count the occurrences of each unigram in the reference and the candidate. Note that the BLEU metric is case-sensitive.

Unigram	m^i_{cand}	m^i_{ref}	$\min(m^i_{cand}, m^i_{ref})$
`the`	3	2	2
`cat`	1	1	1
`is`	0	1	0
`on`	0	1	0
`mat`	1	1	1

The total number of unigrams in the candidate (w^1_t) is 5, so $precision_1$ = (2 + 1 + 1)/5 = 0.8.

Coding Exercise

Train a Machine Translation Model using Seq2Seq

Download data here

Reference:

https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html