

Deep Learning for NLP: TEXT TRANSLATION

Jay Urbain, PhD

Professor, Electrical Engineering and Computer
Science Department

Milwaukee School of Engineering

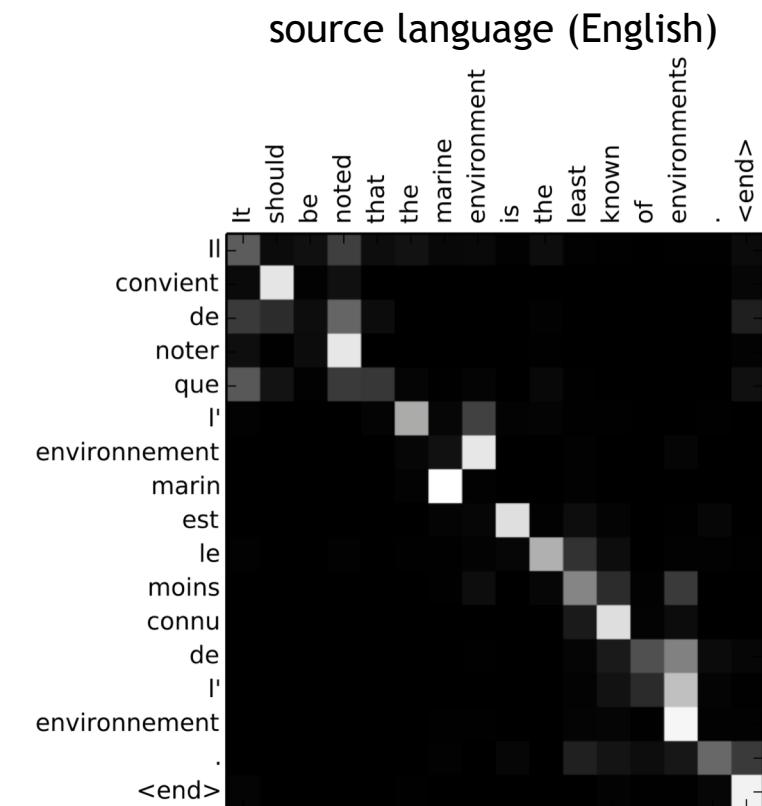
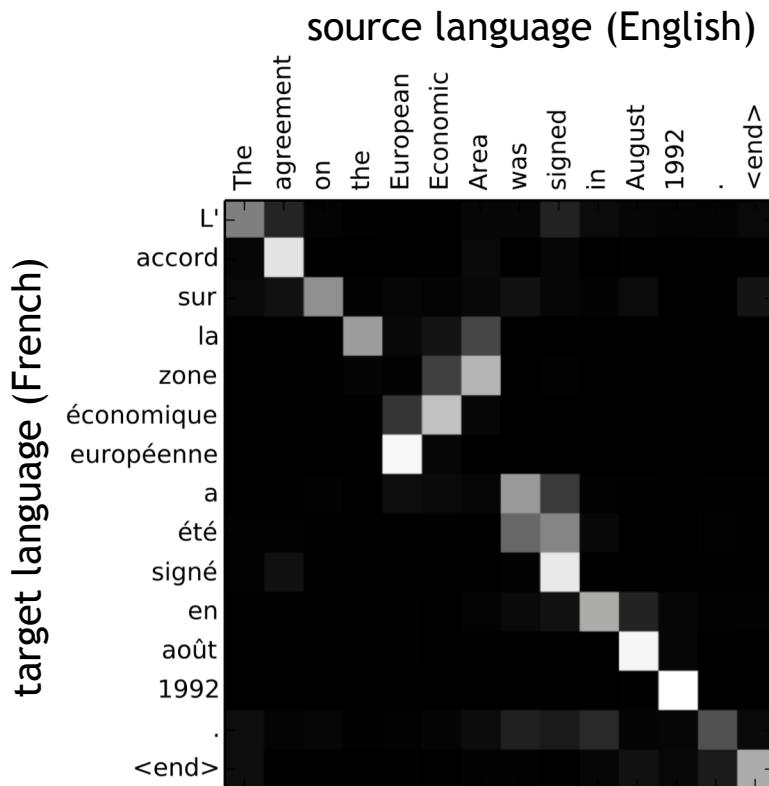
urbain@msoe.edu

July 2018

Neural Machine Translation

- Neural Machine Translation (NMT) is typically used to translate sentences from a source language, e.g. Arabic, to a target language, e.g. English).
- However NMT has a wide range of applications.

Neural Machine Translation

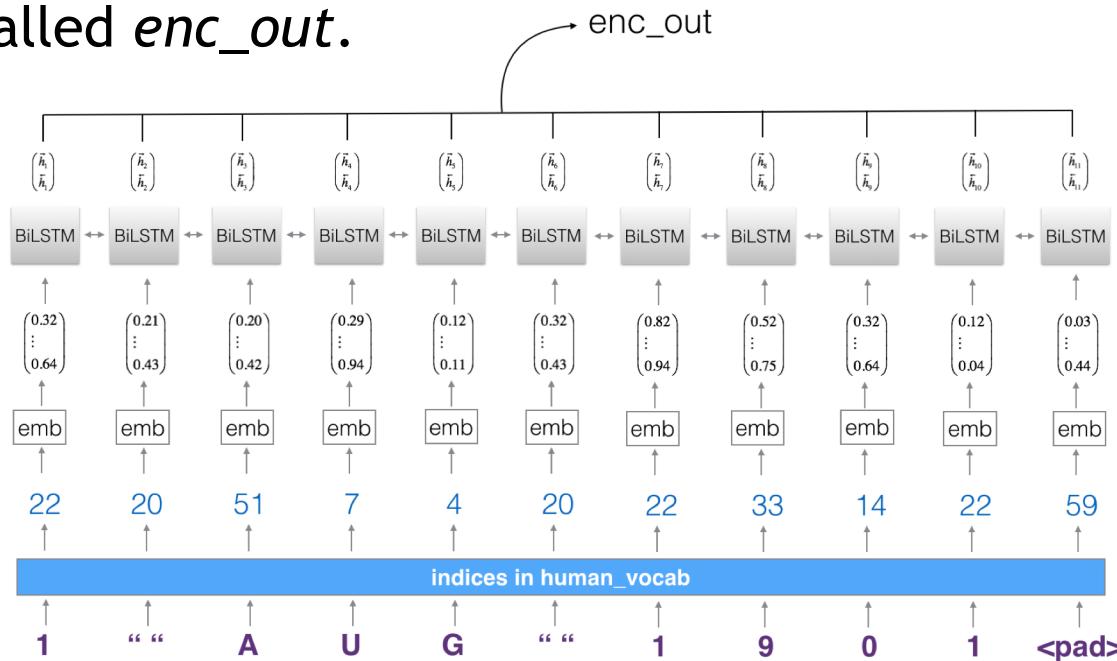


Neural Machine Translation

- We're going build an NMT model to translate human readable dates.
- Example: "the 29th of August 1958", "03/30/1968", "24 JUNE 1987") into machine readable dates (e.g. "1958-08-29", "1968-03-30", "1987-06-24").
- We'll be standardizing the dates on the US date encoding of MM/dd/YYYY when generating the dates.
- The machine readable format is in YYYY-MM-dd format.

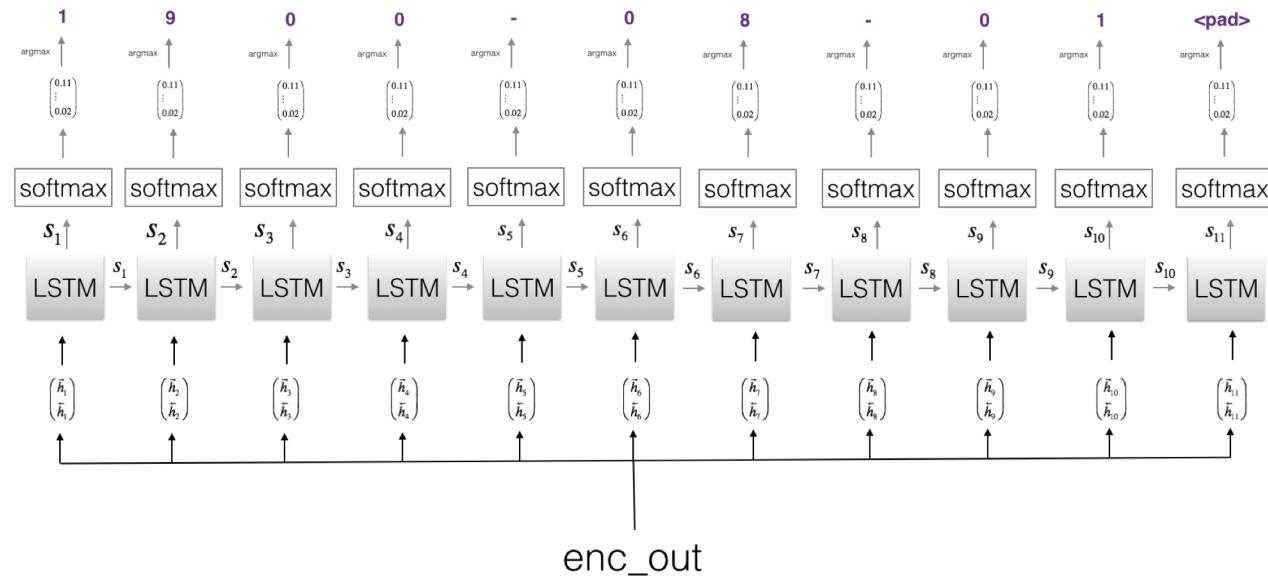
Encoder-Decoder Model: ENCODER

- A preprocessed source date is encoded using a Bi-directional LSTM.
- The sequence of hidden states are returned and stored in an object called *enc_out*.



Encoder-Decoder Model: DECODER

- First, a preprocessed source date is encoded using a Bi-directional LSTM. The sequence of hidden states are returned and stored in an object called *enc_out*. *enc_out* is then given as an input sequence to the *decoder*.



ATTENTION

Intuition:

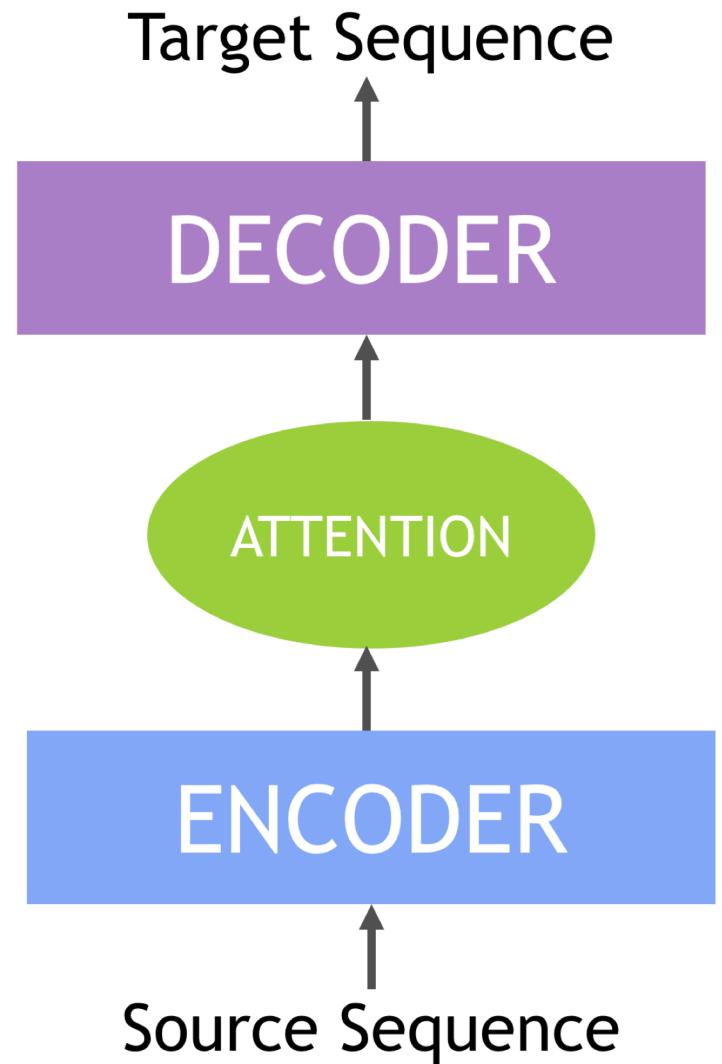
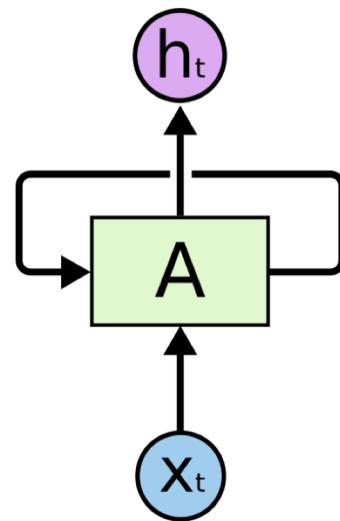
If you had to translate a book's paragraph from French to English, you would not read the whole paragraph, then close the book and translate.

You would first read the paragraph. Then, while translating, you would read and focus on the parts of the paragraph corresponding to the parts you are currently translating.

Attention mechanism is a technique in Deep Learning to help the model drive its focus onto important parts of the input.

NMT Components

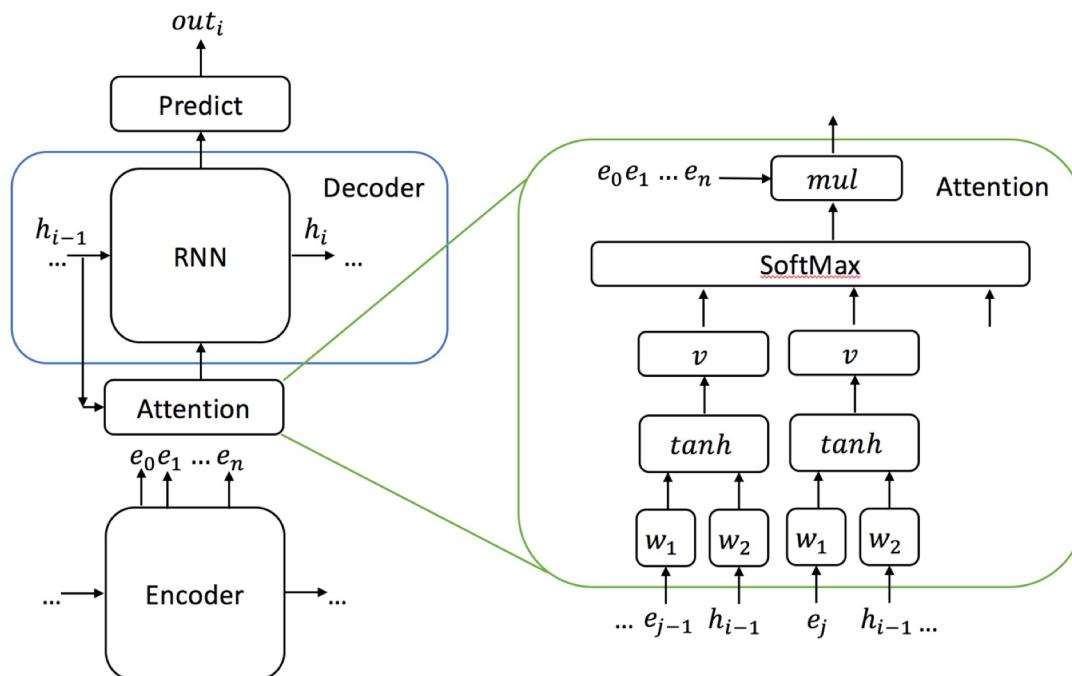
- Encoder
 - Embedding Layer
 - RNN cells
- Decoder
 - RNN cells
 - Embedding Layer
- Attention Layer



ATTENTION

- After running a source date in the encoder (Bi-LSTM), you would like to give all the hidden states to the decoder to get back the target date.
- However, at every step in the decoding process, you'd like your model to be able to tell which hidden state is more important to use.
- So, inserting the attention block between encoder and decoder is the new architecture we will use to decipher the important parts of the input to pay attention to.

- The attention block takes the outputs from the encoder, and applies different weights to them with tanh as the activation function.
- Softmax will choose the most relevant vector and apply attention to that output from the encoder.



LAB DISCUSSION / OVERVIEW

LAB PROCESS

1. Build and evaluate the simple NMT, then NMT with attention:
2. Visual network architecture
3. Train model
4. Compare test, validation and training accuracy
5. Attempt to improve the model
 - Options:
 1. Beam search

ATTENTION MAPS IN TRANSLATION

