Contextualized Embeddings with Recurrent Architectures

Motivation

- Static word embeddings cannot account for polysemy
- Contextualized embeddings: Representations of words in context
- most popular 2012-2018

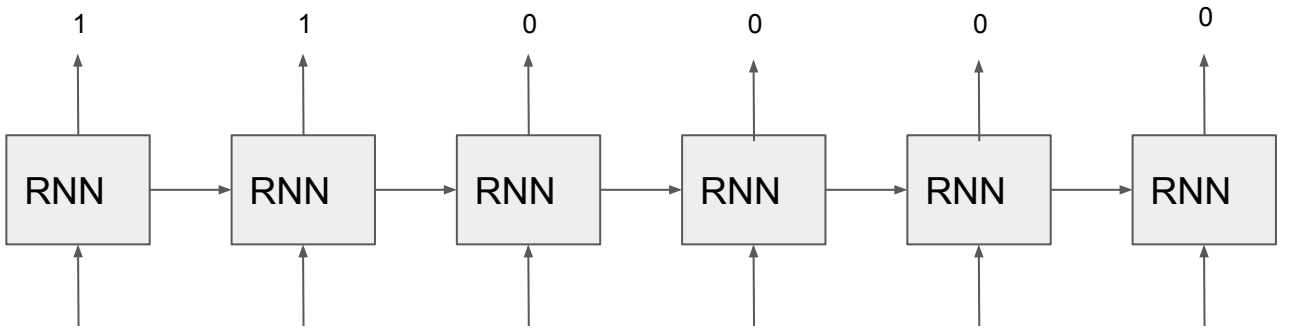
grass on the river bank

vs.

He robbed the bank and fled on a bicycle.

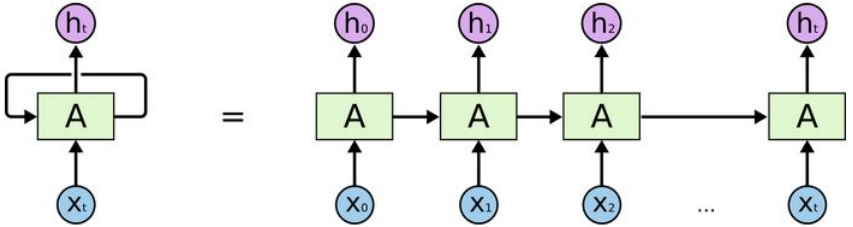
Static Embeddings and Recurrent architectures

- Example: NER with static embeddings & RNN

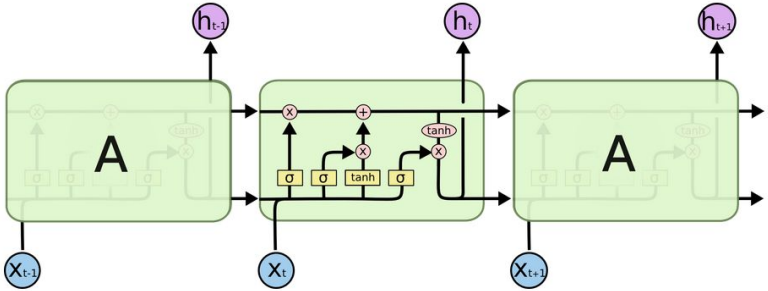


Recurrent Architectures

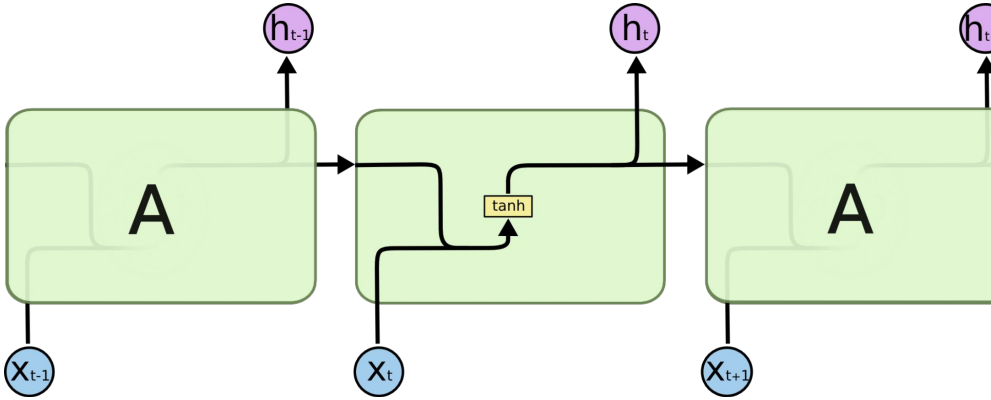
RNNs and LSTM



An unrolled recurrent neural network.



The repeating module in an LSTM contains four interacting layers.

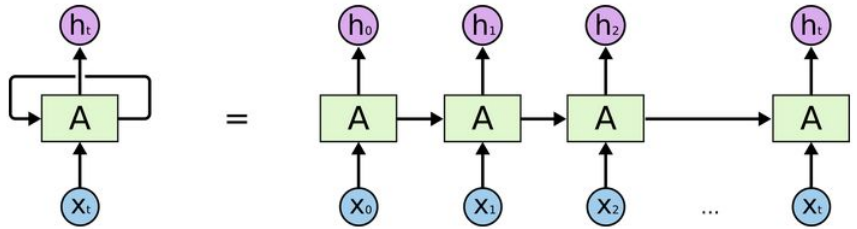


Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Pre-trained contextualization or not?

Two usage possibilities

- trained word vectors as input, RNN is trained for every application
- release fully-trained neural network with all parameters (e.g. ELMo, Peters et al. 2018)



An unrolled recurrent neural network.

ELMo (Peters et al. 2018)

- ELMO = “Embeddings from Language Models”
- bidirectional LSTM architecture
- pretrained on LM task, original paper reports results on 6 applications