SEQUENCE TO SEQUENCE LEARNING WITH RECURRENT NEURAL NETWORKS

Aditi Nair

August 26, 2016

Questions:

1. What is (neural) language modelling?
2. Why do we care about Recurrent Neural Networks (RNNs) and the Long Short Term Memory (LSTM) architecture?
3. What exactly are RNNs and LSTMs and how do they encode information about language and sequences?
4. How can we use these tools for bots?

What is (Neural) Language Modelling?

GOAL: What is the probability of a sentence $S = (x_1, ..., x_n)$?

- In math, what is:

$$p(x_1, x_2, ..., x_n)$$

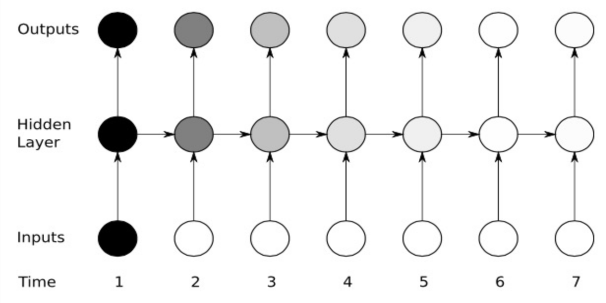
- Recall: $p(x, y) = p(x)p(y|x)$
- So

$$\begin{aligned} p(S) &= p(x_1) \cdot p(x_2|x_1) \cdot ... \cdot p(x_n|x_{n-1}, ...x_2, x_1) \\ &= \prod_{i=1}^{n} p(x_i|x_{<i}) \end{aligned}$$

Why do we care about RNNs and LSTMs?

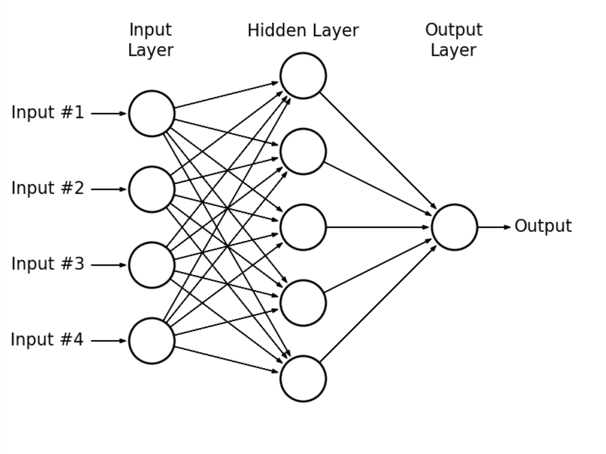
- RNNs and LSTMs make language modelling ”easier”
  - No more manual feature selection!
- And better!
  - Sutskever et. Al (2014) used LSTMs to achieve a BLEU score of 34.81 on a standard English to French translation tasks
  - In comparison, a baseline phrase-based statistical machine translation achieved a BLEU score of 33.3 on the same task

Recurrent Neural Networks



Source: http://eric-yuan.me/wp-content/uploads/2015/06/1.jpg

(Classic) Feed-Forward Neural Network



Source: https://www.cs.bgu.ac.il/ ben-shahar/Teaching/Computational-Vision/StudentProjects/ICBV061/ICBV-2006-1-TorIvry-ShaharMichal/Images/network.gif

Recurrent Neural Networks, contd.

- RNNs are *parametric, recursive* functions.
- Start with a sequence $x_1, x_2, ...x_T$
- Initialize a vector $h_0 \in \mathbb{R}^d$
- For $t = 1, 2, ...T$:
    - Define $h_t = f(x_t, h_{t-1})$
- Entire sequence is summarized by $h_T \in \mathbb{R}^d$

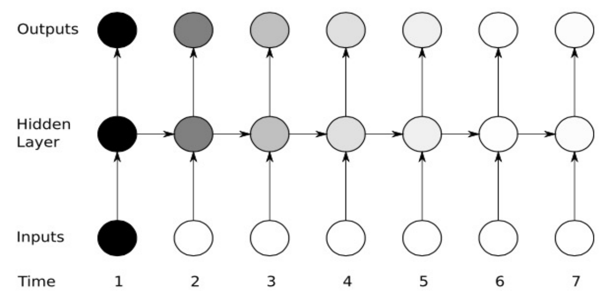
What does f look like?

- $f(x_t, h_{t-1}) = g(W\phi(x_t) + Uh_{t-1})$
- Recall $h_{t-1} \in \mathbb{R}^d$
- $\phi(x_t)$ transforms x_t into an element of $\mathbb{R}^n$
- $W \in \mathbb{R}^{d \times n}$
- $U \in \mathbb{R}^{d \times d}$
- g is the "activation" function: *tanh*

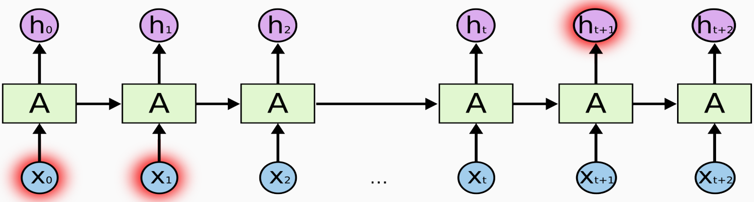
What does $\phi(x_t)$ look like?

- It depends...
- Typically: use a one-hot encoding to transform x_t to a vector in $\mathbb{R}^{|\mathbb{V}|}$
  - Call this $\tilde{x}_t$
- Then define $\phi(x_t) = E\tilde{x}_t$
  - $E \in \mathbb{R}^{n \times |V|}$
- Tune choice of E and skip feature engineering!

Are RNNs good enough?



The Long-term Dependency Problem



"I grew up in France… I speak fluent *French*."

Gated Recurrent Units (GRU)

- Problem: At each time t we look at all of h_{t-1} to compute $\tilde{h}_t$
  - Use a reset gate $r \in [0, 1]$:

$$\tilde{h}_t = f(x_t, r \odot h_{t-1}) = g(W\phi(x_t) + U(r \odot h_{t-1}))$$

- Problem: Entire history vector is updated at each time t.
  - Use an update gate $u \in [0, 1]$:

$$h_t = (1 - u) \odot h_{t-1} + u \odot \tilde{h}_t$$

Long Short-term Memory (LSTM)

- Similar to GRU - with strong empirical results
- Solves memory problem by separating the output vector h_t from the internal memory vector c_t:
- Memory state:
  - f is the forget gate
  - i is the input gate

$$c_t = f \odot c_{t-1} + i \odot \tilde{c}_t$$

$$\tilde{c}_t = tanh(W_c \phi(x_t) + U_c h_{t-1})$$

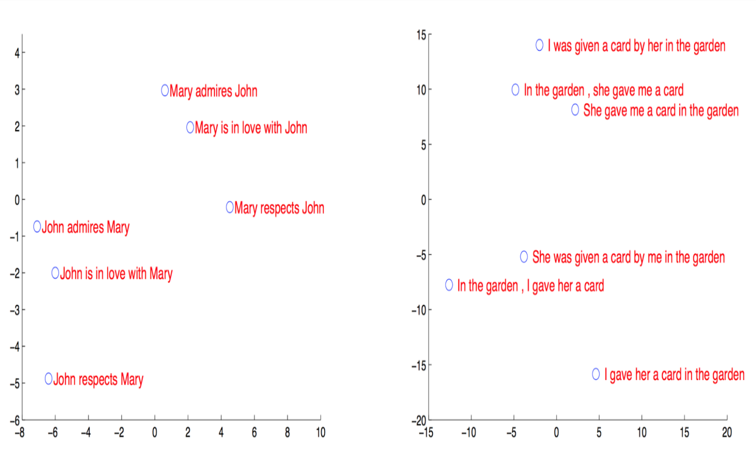
- Output gate:

$$o = tanh(W_o \phi(x_t) + U_o h_{t-1})$$

- Output vector:

$$h_t = o \odot tanh(c_t)$$

Seq2Seq Results



The Seq2Seq Framework

- Translate $(x_1, x_2, ...x_T)$ to $(y_1, ...y_{T'})$
- In math: choose the most likely translation $(y_1, ...y_{T'})$
  - $p(y_1, ...y_{T'}|x_1, x_2...x_T)$
- This is straightforward:
  - Pass $(x_1, ...x_T)$ through an LSTM
  - Define $c = h_T$
  - Now: $p(y_1, ...y_{T'}|x_1, x_2...x_T) = \prod_{t=1}^{T} p(y_t|c, y_{<t})$
  - Use a second LSTM to estimate $p(y_t|c, y_{<t})$
    - For each word w compute:

$$score_w = w^T h_t + b$$

    - Soft-max to turn probabilities into scores

RNN and Bots!

- Similar to MT but…for each emission use the conversation thus far as the input sequence
- Use corpus of past conversations to tune parameters for LSTM
- Example from ”A Neural Conversational Model” (Oriol Vinyals and Quoc Le at Google, 2015)

RNN and Bots!

Human: is morality and ethics the same ?
Machine: yes , sir .
Human: what are the things that i do to be immoral ?
Machine: i don ’t know .
Human: give me some examples of moral actions ...
Machine: i ’m not a moralist .
...
...
Human: be moral !
Machine: be a man !
Human: i really like our discussion on morality and ethics ...
Machine: and how i ’m not in the mood for a philosophical debate .

RNN and Bots!

A Funny Movie.

The End!