Some fancy title: followed by some more text

Alyssa P. Hacker [1] Ben Bitdiddle [2] Lem E. Tweakit [2]

[1]Some Institute [2]Another Institute

Introduction

Recurrent neural networks (RNN) constitute a family of neural network architectures specialized to process sequential data. They do so by leveraging the simple idea of sharing parameters across the model[1]. They have been used in diverse domains for generating sequences such as music and text. They can be trained by processing real data sequences one step at a time and predicting what comes next. In practice, early RNNs designs are unable to store information about far past inputs. This fact diminishes their efficiency at modeling long structure. If the network's predictions are based only on the last few inputs, which are themselves predicted by the networks, then it has less chance to recover from past mistakes. The solution to this problem seems to be a better memory and especially a long-term memory for the network.

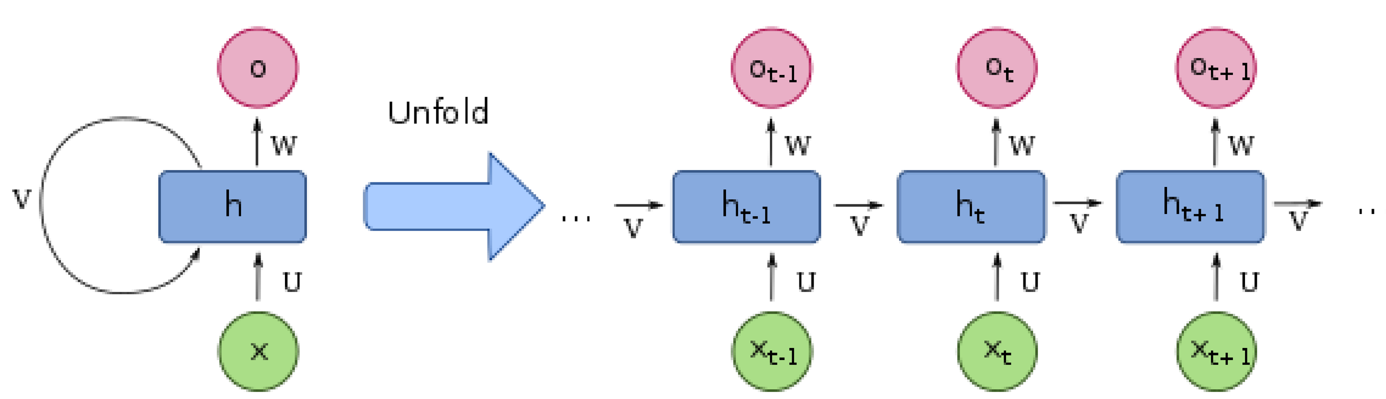
Long Short-Term Memory (LSTM) are RNN designed to solve this problem. They are better at storing informations than standard RNNs. It is important to note that LSTM gives state-of-the-art results in a variety of sequence processing tasks. This is the main reason why we decided to implement a LSTM for our text generation project. We worked on 4 differents dataset: Harry Potter's books, Lord of the ring's books, random quotes and a text from Shakespeare.

RNN structure

RNN comes from the following question: is there a neural network that depends on the full previous context that would model:

$$P(o_1, \ldots, o_T) = \prod_{t=1}^{T} P(o_t | o_1, \ldots o_{t-1})$$

They are feedforward neural networks with the addition of time dependency in the model by introducing edges that span the adjacent time steps in the network. At a given time, the nodes with recurrent edges receive input from the current data and from the output of the hidden layer in the previous state, see figure below. Thus, an input at time t can influence the output at time $t+\delta$.



RNN parameter udpates and vanishing-exploding gradient

$$h_t = \sigma(Ux_t + Vh_{(t-1)} + b_h),$$
$$o_t = \text{softmax}(Wh_t + b_o)$$

Here, U, V and W are weights matrix. The vectors b are bias parameters. Learning with RNN is challenging due to dependencies between long time steps. Consider the gradient with respect to h_t of $o_{t+\delta}$. How does it vary with δ? Following the graph above and applying the chain rule we can see that

$$\nabla_{h_t} o_{t+\delta} = \left(\prod_{k=t+1}^{t+\delta} V^T \text{diag}(1 - h_k^2) \right) \nabla_{h_{t+\delta}} o_{t+\delta}.$$

Thus, as δ grows, the gradient grows exponentially with V. If V is small or large than the gradient will either vanish or explode. This problem is well known. Solutions exist, which brings us to present the LSTM.

Fusce aliquam magna velit

Et rutrum ex euismod vel. Pellentesque ultricies, velit in fermentum vestibulum, lectus nisi pretium nibh, sit amet aliquam lectus augue vel velit. Suspendisse rhoncus massa porttitor augue feugiat molestie. Sed molestie ut orci nec malesuada. Sed ultricies feugiat est fringilla posuere.

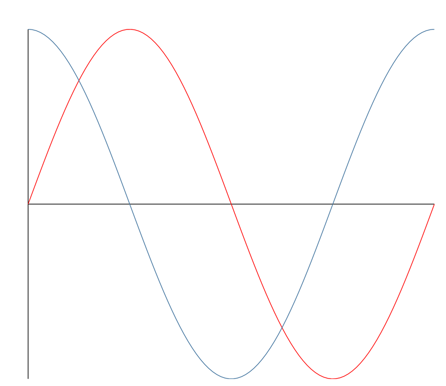


Figure: Another figure caption.

Nam cursus consequat egestas

Nulla eget sem quam. Ut aliquam volutpat nisi vestibulum convallis. Nunc a lectus et eros facilisis hendrerit eu non urna. Interdum et malesuada fames ac ante *ipsum primis* in faucibus. Etiam sit amet velit eget sem euismod tristique. Praesent enim erat, porta vel mattis sed, pharetra sed ipsum. Morbi commodo condimentum massa, *tempus venenatis* massa hendrerit quis. Maecenas sed porta est. Praesent mollis interdum lectus, sit amet sollicitudin risus tincidunt non.

Etiam sit amet tempus lorem, aliquet condimentum velit. Donec et nibh consequat, sagittis ex eget, dictum orci. Etiam quis semper ante. Ut eu mauris purus. Proin nec consectetur ligula. Mauris pretium molestie ullamcorper. Integer nisi neque, aliquet et odio non, sagittis porta justo.

- Sed consequat id ante vel efficitur. Praesent congue massa sed est scelerisque, elementum mollis augue iaculis.
  - In sed est finibus, vulputate nunc gravida, pulvinar lorem. In maximus nunc dolor, sed auctor eros porttitor quis.
  - Fusce ornare dignissim nisi. Nam sit amet risus vel lacus tempor tincidunt eu a arcu.
  - Donec rhoncus vestibulum erat, quis aliquam leo gravida egestas.
- Sed luctus, elit sit amet dictum maximus, diam dolor faucibus purus, sed lobortis justo erat id turpis.
- Pellentesque facilisis dolor in leo bibendum congue. Maecenas congue finibus justo, vitae eleifend urna facilisis at.

A block containing some math

Nullam non est elit. In eu ornare justo. Maecenas porttitor sodales lacus, ut cursus augue sodales ac.

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

Interdum et malesuada fames $\{1, 4, 9, \ldots\}$ ac ante ipsum primis in faucibus. Cras eleifend dolor eu nulla suscipit suscipit. Sed lobortis non felis id vulputate.

A heading inside a block

Praesent consectetur mi $x^2 + y^2$ metus, nec vestibulum justo viverra nec. Proin eget nulla pretium, egestas magna aliquam, mollis neque. Vivamus dictum $\mathbf{u}^\intercal\mathbf{v}$ sagittis odio, vel porta erat congue sed. Maecenas ut dolor quis arcu auctor porttitor.

Another heading inside a block

Sed augue erat, scelerisque a purus ultricies, placerat porttitor neque. Donec $P(y \mid x)$ fermentum consectetur $\nabla_x P(y \mid x)$ sapien sagittis egestas. Duis eget leo euismod nunc viverra imperdiet nec id justo.

Nullam vel erat at velit convallis laoreet

Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Phasellus libero enim, gravida sed erat sit amet, scelerisque congue diam. Fusce dapibus dui ut augue pulvinar iaculis.

First column	Second column	Third column	Fourth
Foo	13.37	384,394	α
Bar	2.17	1,392	β
Baz	3.14	83,742	δ
Qux	7.59	974	γ

Table: A table caption.

Donec quis posuere ligula. Nunc feugiat elit a mi malesuada consequat. Sed imperdiet augue ac nibh aliquet tristique. Aenean eu tortor vulputate, eleifend lorem in, dictum urna. Proin auctor ante in augue tincidunt tempor. Proin pellentesque vulputate odio, ac gravida nulla posuere efficitur. Aenean at velit vel dolor blandit molestie. Mauris laoreet commodo quam, non luctus nibh ullamcorper in. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Nulla varius finibus volutpat. Mauris molestie lorem tincidunt, iaculis libero at, gravida ante. Phasellus at felis eu neque suscipit suscipit. Integer ullamcorper, dui nec pretium ornare, urna dolor consequat libero, in feugiat elit lorem euismod lacus. Pellentesque sit amet dolor mollis, auctor urna non, tempus sem.

References