LSTM

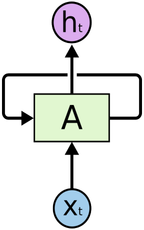
Golubev Kirill

March 29, 2019

heavily based on:
colah blogpost

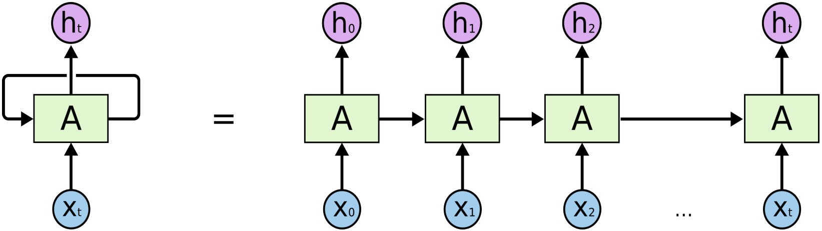
RNN review

concept



RNN review

concept

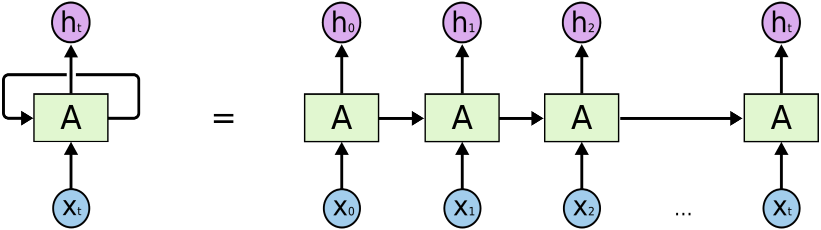


most important thing:

?

RNN review

concept

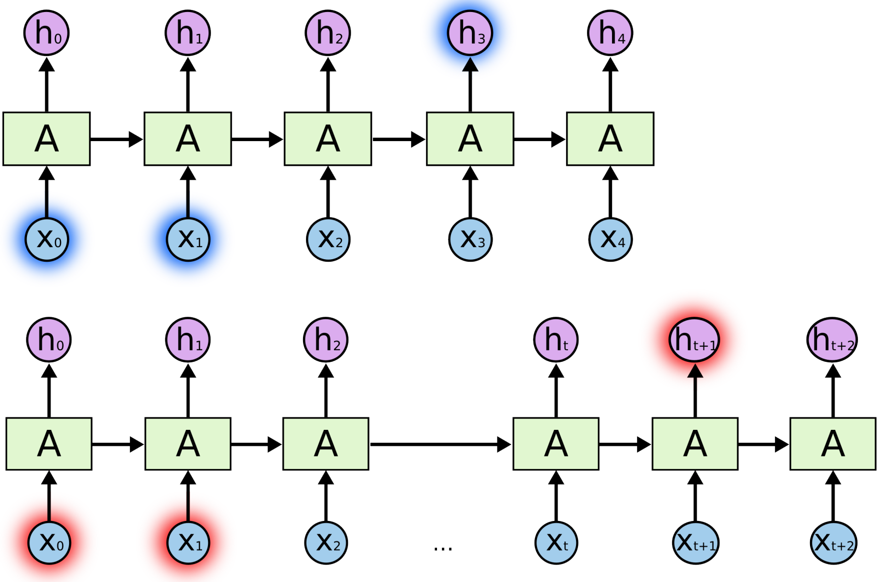


most important thing:

It is just another input dimension

RNN review

problems



RNN review

Functioning example

There are clouds in the ****.

Malfunctioning example

She lived in France in her childhood. .... She fluently speaks ****.

RNN review

Functioning example

There are clouds in the ****.

Malfunctioning example

She lived in France in her childhood. .... She fluently speaks ****.

Main problems:
Long term dependency.

Fantasy:

dream recurrent cell

Fantasy:

dream recurrent cell

What do we have:

$$
\begin{aligned}
s_{t-1} &= (s_1, s_2, ..., s_n) - \textit{previous state} \\
h_{t-1} &= (h_1, h_2, ..., h_q) - \textit{previous output} \\
x_{t-1} &= (x_1, x_2, ..., x_p) - \textit{current input}
\end{aligned}
$$

What do we want

What to forget:

What to remember:

What to output:

Fantasy:

What do we have:

$$s_{t-1} = (s_1, s_2, ..., s_n) - \textit{previous state}$$
$$h_{t-1} = (h_1, h_2, ..., h_q) - \textit{previous output}$$
$$x_{t-1} = (x_1, x_2, ..., x_p) - \textit{current input}$$

What do we want

What to forget: $\tilde{s}_{t-1} = f(x_t, h_{t-1}) \cdot s_{t-1}, F \in [0, 1]^n$

What to remember:

What to output:

Fantasy:

dream recurrent cell

What do we have:

$$s_{t-1} = (s_1, s_2, ..., s_n) - \textit{previous state}$$
$$h_{t-1} = (h_1, h_2, ..., h_q) - \textit{previous output}$$
$$x_{t-1} = (x_1, x_2, ..., x_p) - \textit{current input}$$

What do we want:

What to forget: $\tilde{s}_{t-1} = f(x_t, h_{t-1}) \cdot s_{t-1}, F \in [0,1]^n$

What to remember: $\tilde{s}_t = r(x_t, h_{t-1}) \cdot m(x_t, h_{t-1})$

$r \in [0,1]^n, \ \ m \in [-1,1]^n$

new state: $s_t = \tilde{s}_{t-1} + \tilde{s}_t$

What to output:

Fantasy:

dream recurrent cell

What do we have:

$$s_{t-1} = (s_1, s_2, ..., s_n) - \textit{previous state}$$
$$h_{t-1} = (h_1, h_2, ..., h_q) - \textit{previous output}$$
$$x_{t-1} = (x_1, x_2, ..., x_p) - \textit{current input}$$

What do we want:

What to forget: $\tilde{s}_{t-1} = f(x_t, h_{t-1}) \cdot s_{t-1}, F \in [0, 1]^n$

What to remember: $\tilde{s}_t = r(x_t, h_{t-1}) \cdot m(x_t, h_{t-1})$

$r \in [0, 1]^n, \;\; m \in [-1, 1]^n$

new state: $s_t = \tilde{s}_{t-1} + \tilde{s}_t$

What to output: $h_t = act(W \cdot (x_t, h_{t-1}) + b) \cdot i(s_t)$

Fantasy:

dream recurrent cell

What do we want:

What to forget: $\tilde{s}_{t-1} = f(x_t, h_{t-1}) \cdot s_{t-1}, F \in [0,1]^n$

What to remember: $\tilde{s}_t = r(x_t, h_{t-1}) \cdot m(x_t, h_{t-1})$

$r \in [0,1]^n, \ \ m \in [-1,1]^n$

new state: $s_t = \tilde{s}_{t-1} + \tilde{s}_t$

What to output: $h_t = act(W \cdot (x_t, h_{t-1}) + b) \cdot i(s_t)$

What we do not have:

$f \in [0,1]^n$ – *forgetting function*

$r \in [0,1]^n, \ \ m \in [-1,1]^n$, *remembering function and memories itself*

$i \in [-1,1]^n$ – *impact function*

What we do not have:

$f \in [0,1]^n$ – *forgetting function*

$r \in [0,1]^n$, $m \in [-1,1]^n$, *remembering function and memories itself*

$i \in [-1,1]^n$ – *impactFunction*

Let's train them!

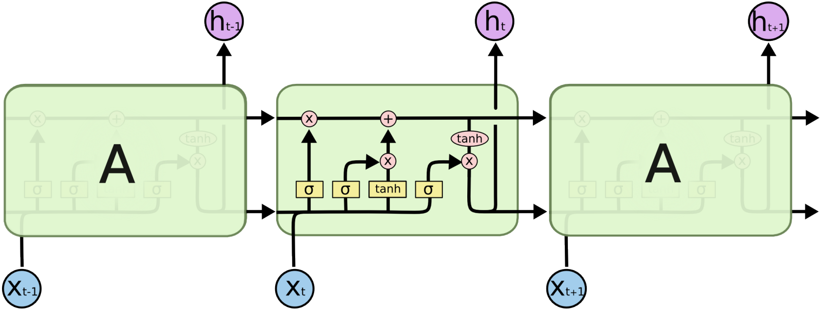
What we can have:

$$f \in [0,1]^n F = \sigma(W_f \cdot (x_t, h_{t-1}) + b_f)$$
$$r \in [0,1]^n, \;\; m \in [-1,1]^n,$$

$$r = \sigma(W_r \cdot (x_t, h_{t-1}) + b_r),$$
$$m = tanh(W_m \cdot (x_t, h_{t-1}) + b_m)$$

$$i \in [-1,1]^n, \;\; i = tanh(s_t)$$

LSTM



LSTM

Peephole variation

$$f \in [0,1]^n F = \sigma(W_f \cdot (\mathbf{s_{t-1}}, x_t, h_{t-1}) + b_f)$$
$$r \in [0,1]^n,\ \ m \in [-1,1]^n,$$

$$r = \sigma(W_r \cdot (\mathbf{s_{t-1}}, x_t, h_{t-1}) + b_r),$$
$$m = tanh(W_m \cdot (x_t, h_{t-1}) + b_m)$$

$$i \in [-1,1]^n,\ \ i = tanh(s_t)$$
$$h_t = act(W \cdot (\mathbf{s_t}, x_t, h_{t-1}) + b) \cdot i(s_t)$$

LSTM

Another one

$$f \in [0,1]^n F = \sigma(W_f \cdot (x_t, h_{t-1}) + b_f)$$
$$r \in [0,1]^n, \;\; m \in [-1,1]^n,$$

$$s_t = \tilde{s}_{t-1} + \tilde{s}_t \Leftrightarrow s_t = f \cdot s_{t-1} + r \cdot m(x_t, h_{t-1})$$
$$r = 1 - f$$
$$s_t = \tilde{s}_{t-1} + \tilde{s}_t \Leftrightarrow s_t = f \cdot s_{t-1} + (1 - f) \cdot m(x_t, h_{t-1})$$

GRU

gated recurrent unit

Let's make LSTM little bit simpler by removing it's state.
It's role will be performed by previous output.

$$z = \sigma(W_z \cdot (x_t, h_{t-1})),\ z \in [0,1]^q$$
$$r = \sigma(W_r \cdot (x_t, h_{t-1})),\ r \in [0,1]^q$$
$$\tilde{h_t} = tanh(W \cdot (x_t, r \cdot h_{t-1}))$$
$$h_t = (1 - z) \cdot h_{t-1} + z * \tilde{h_t}$$

GRU

gated recurrent unit

$$z = \sigma(W_z \cdot (x_t, h_{t-1})),\ \ z \in [0,1]^q$$
$$r = \sigma(W_r \cdot (x_t, h_{t-1})),\ \ r \in [0,1]^q$$
$$\tilde{h}_t = tanh(W \cdot (x_t, r \cdot h_{t-1}))$$
$$h_t = (1 - z) \cdot h_{t-1} + z * \tilde{h}_t$$