5. Gating and LSTM

Gating and LSTM

exploding

gradient issues, so do recurrent neural networks.

Specific architectures such as the LSTM recurrent neural

network maintain a better control

over the information that's retained,

updated along the sequence, and they are therefore

easier to train to do what we want them to do.

End of transcript. Skip to the start.

9:21 / 9:21 1.0x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

xuetangX.com 学堂在线

Gating

1/1 point (graded)
Recall that the most simple, single-layered RNN can be written in equation as:

$$s_t = \tanh\left(W^{s,s} s_{t-1} + W^{s,x} x_t\right).$$

Recognize that, in the above formulation, s_t is always overwritten with the calculated result $\tanh\left(W^{s,s} s_{t-1} + W^{s,x} x_t\right)$.

Now, we introduce a gate vector g_t of the same dimension as s_t, which determines "how much information to overwrite in the next state." In equation, a single-layered gated RNN can be written as:

$$\begin{aligned} g_t &= \text{sigmoid}\left(W^{g,s} s_{t-1} + W^{g,x} x_t\right) \\ s_t &= (1 - g_t) \odot s_{t-1} + g_t \odot \tanh\left(W^{s,s} s_{t-1} + W^{s,x} x_t\right). \end{aligned}$$

where the sign $\odot$ denotes element-wise multiplication. Now, which of the following is true about the gate g_t? (Choose all those apply.)

- [ ] If the ith element of g_t is 1, the ith element of s_t and that of s_{t-1} are equal
- [x] If the ith element of g_t is 0, the ith element of s_t and that of s_{t-1} are equal ✔

- [ ] If g_t is a vector whose elements are all 1, s_t and s_{t-1} are equal
- [x] If g_t is a vector whose elements are all 0, s_t and s_{t-1} are equal ✔

✔

Solution:

Let the ith element of s_t,g_t,s_{t-1} be s_t^i,g_t^i,s_{t-1}^i.

If the ith element of g_t is 0, $(1 - g_t^i) = 1$, so

$$s_t^i \; = s_{t-1}^i.$$

Thus, if the ith element of g_t is 0, the ith element of s_t and that of s_{t-1} are equal. Also, if g_t is a vector whose elements are all 0, s_t and s_{t-1} are equal.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

LSTM

1/1 point (graded)
Which of the following components of an LSTM represent the context or state? (Choose all that apply.)

- [x] c_t ✔
- [x] h_t ✔
- [ ] o_t
- [ ] i_t

✔

Solution:

c_t represents the memory cell, and h_t represents the visible state. Together they make up the context or state. The other two choices are the output and input gate, respectively. They simply accomodate new inputs and output predictions, and are not part of the context/state

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

LSTM Calculations

1/1 point (graded)
Let all the neural network's weight matrices, the hidden state, and the memory cell be a scalar 1. Let the new x-value be 5. Calculate the value of the new hidden state. Round sigmoid to 1 or 0, and round $\tanh$ to -1 or 1.

$$f_t = \text{sigmoid}(W^{f,h}h_{t-1} + W^{f,x}x_t) \quad \text{forget gate}$$
$$i_t = \text{sigmoid}(W^{i,h}h_{t-1} + W^{i,x}x_t) \quad \text{input gate}$$
$$o_t = \text{sigmoid}(W^{o,h}h_{t-1} + W^{o,x}x_t) \quad \text{output gate}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W^{c,h}h_{t-1} + W^{c,x}x_t) \quad \text{memory cell}$$
$$h_t = o_t \odot \tanh(c_t) \quad \text{visible state}$$

- -1
- 1 ✔
- 0
- 5

Solution:

The forget gate is equal to sigmoid(6), or 1. The same applies for the input and output gate. The memory cell is equal to 1+ tanh(1+6), which is 2. The new hidden state is therefore tanh(2), or 1.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 10. Recurrent Neural Networks 1 / 5. Gating and LSTM

Show Discussion