

Chapter20-2 Recurrent Neural Network

1. RNN
 - a) Function: $h', y = f(h, x)$
 - b) No matter how long the input/output sequence is, we only need one function
2. Deep RNN
3. Bidirectional RNN
 - a) $a', p = f_1(a, x)$
 - b) $b', q = f_2(b, x)$
 - c) $y = f_3(p, q)$
4. Naïve RNN
 - a) $h' = \sigma(W^h \cdot h + W^i \cdot x)$
 - b) $y = \sigma(W^o \cdot h')$
5. LSTM
 - a) Parameters
 - i) c changes slowly: c' is c added by something
 - ii) h changes faster: h' and h can be very different
 - b) Mechanism
 - i) $z = \tanh(W \cdot \text{concatenated}(x^t, h^{t-1}))$
 - ii) $z^i = \sigma(W^i \cdot \text{concatenated}(x^t, h^{t-1}))$
 - iii) $z^f = \sigma(W^f \cdot \text{concatenated}(x^t, h^{t-1}))$
 - iv) $z^o = \sigma(W^o \cdot \text{concatenated}(x^t, h^{t-1}))$
 - v) $c^t = z^i \cdot z + z^f \cdot c^{t-1}$
 - vi) $h^t = z^o \cdot \tanh(c^t)$
 - vii) $y^t = \sigma(W' \cdot h^t)$
6. Gated Recurrent Unit
 - a) Basically has the same performance with LSTM
 - b) Less parameters
7. Performance
 - a) Forget gate and output gate activation function is critical
 - b) Gates Importance: forget > input > output
 - c) Large bias for forget gate is helpful
8. More Application
 - a) Many to one
 - i) Sentiment Analysis
 - ii) Key Term Extraction
 - b) Many to many (Output is shorter)
 - i) Speech Recognition (Connectionist Temporal Classification)
 - c) Sequence to sequence
 - i) Machine translation
 - d) Attention-based model
 - i) Reading Head Controller
 - ii) Writing Head Controller