LSTM: A Search Space Odysseys

Muşat Bogdan-Adrian

February 2017

Recurrent Neural Networks - RNNs

- Used to deal with sequential data, where there is a temporal dependence from a time instance to another
- Mathematically, they model a conditional distribution of the form $P(x_t|x_{t-1}, ..., x_2, x_1)$, where x_t is the current input at time t
- The output of a vanilla RNN cell at each time step is computed using the current input x_t and also the previous cell state h_{t-1}[1] as:
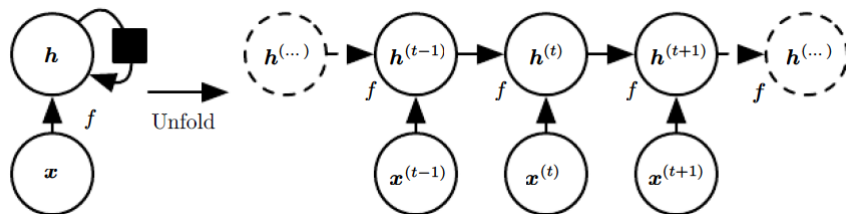
$$y_t = tanh(Ux_t + Wh_{t-1} + b),$$

where $U \in R^{M \times N}$, $W \in R^{N \times N}$ are the shared parameter matrices for the RNN cells and $b \in R^N$ is the bias

---

[1]Meant to encompass a summary of the past information

Recurrent Neural Networks Unrolled

(a) RNN unrolled[2]

[2]Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.