Hierarchical Multiscale Recurrent Neural Networks

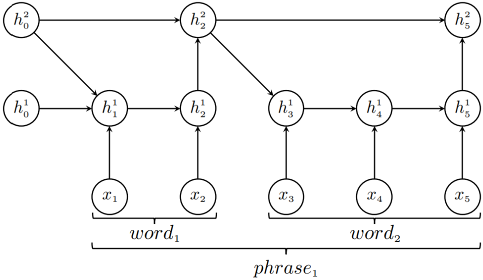
Chung et al., ICLR 2017
Speaker : Semin Choi

Department of Statistics, Seoul National University, South Korea

April 26, 2018

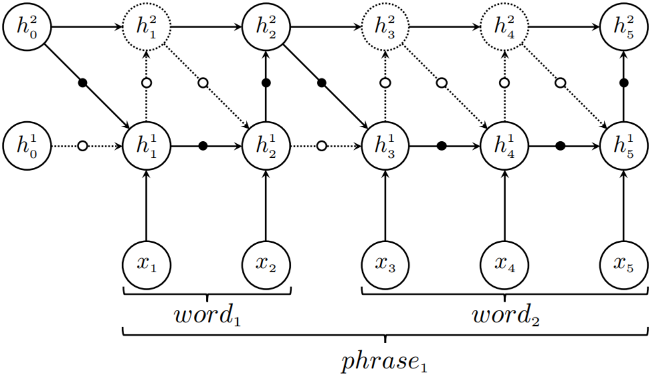
Motivation

- The Hierarchical RNN architecture



- It requires the knowledge of the hierarchical boundaries.
- Can an RNN discover such hierarchical multiscale structure without explicit hierarchical boundary information?
- Ex) handwriting sequence generation

Hierarchical Multiscale LSTM



Hierarchical Multiscale LSTM

- A key element of the proposed model is the introduction of a parameterized boundary detector.
- Consider an HM-LSTM model of L layers ($l = 1, ..., L$) which, at each layer l, performs the following update at time step t:

$$\mathbf{h}_t^l, \mathbf{c}_t^l, z_t^l = f_{\text{HM-LSTM}}^l(\mathbf{c}_{t-1}^l, \mathbf{h}_{t-1}^l, \mathbf{h}_t^{l-1}, \mathbf{h}_{t-1}^{l+1}, z_{t-1}^l, z_t^{l-1}).$$

  Here, $\mathbf{h}$ and $\mathbf{c}$ denote the hidden and cell states, respectively.
- The function $f_{\text{HM-LSTM}}^l$ is implemented as follows:

$$\mathbf{c}_t^l = \begin{cases} \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l + \mathbf{i}_t^l \odot \mathbf{g}_t^l & \text{if } z_{t-1}^l = 0 \text{ and } z_t^{l-1} = 1 \text{ (UPDATE)} \\ \mathbf{c}_{t-1}^l & \text{if } z_{t-1}^l = 0 \text{ and } z_t^{l-1} = 0 \text{ (COPY)} \\ \mathbf{i}_t^l \odot \mathbf{g}_t^l & \text{if } z_{t-1}^l = 1 \text{ (FLUSH)}, \end{cases}$$

$$\mathbf{h}_t^l = \begin{cases} \mathbf{h}_{t-1}^l & \text{if COPY} \\ \mathbf{o}_t^l \odot \tanh(\mathbf{c}_t^l) & \text{otherwise.} \end{cases}$$

  Here, $(\mathbf{f}, \mathbf{i}, \mathbf{o})$ are forget, input, output gates, and $\mathbf{g}$ is a cell proposal vector.

Hierarchical Multiscale LSTM

$$\begin{pmatrix} \mathbf{f}_t^l \\ \mathbf{i}_t^l \\ \mathbf{o}_t^l \\ \mathbf{g}_t^l \\ \tilde{z}_t^l \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \text{tanh} \\ \text{hard sigm} \end{pmatrix} f_{\text{slice}}\left(\mathbf{s}_t^{\text{recurrent}(l)} + \mathbf{s}_t^{\text{top-down}(l)} + \mathbf{s}_t^{\text{bottom-up}(l)} + \mathbf{b}^{(l)}\right)$$

where

$$\begin{aligned} \mathbf{s}_t^{\text{recurrent}(l)} &= U_l^l \mathbf{h}_{t-1}^l, \\ \mathbf{s}_t^{\text{top-down}(l)} &= z_{t-1}^l U_{l+1}^l \mathbf{h}_{t-1}^{l+1}, \\ \mathbf{s}_t^{\text{bottom-up}(l)} &= z_t^{l-1} W_{l-1}^l \mathbf{h}_t^{l-1}, \\ \text{hard sigm}(x) &= \max(0, \min(1, \frac{ax+1}{2})). \end{aligned}$$

Here, we use $W_{l-1}^l \in \mathbb{R}^{(4dim(\mathbf{h}^l)+1)\times dim(\mathbf{h}^{l-1})}$, $U_l^l \in \mathbb{R}^{(4dim(\mathbf{h}^l)+1)\times dim(\mathbf{h}^l)}$, $U_{l+1}^l \in \mathbb{R}^{(4dim(\mathbf{h}^l)+1)\times dim(\mathbf{h}^{l+1})}$ and $\mathbf{b} \in \mathbb{R}^{4dim(\mathbf{h}^l)+1}$.

Hierarchical Multiscale LSTM

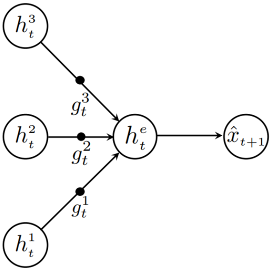
Finally, the binary boundary state z_t^l is obtained by:

$$z_t^l = f_{bound}(\tilde{z}_t^l)$$

we can either use a deterministic step function or sample from a Bernoulli distribution for f_{bound}.

Character-Level Language Modelling

- The output module when $L = 3$.



$$g_t^l = \mathrm{sigm}(\mathbf{w}^l[\mathbf{h}_t^1; \ldots; \mathbf{h}_t^L])$$

- The output embedding $\mathbf{h}_t^e$ is computed by :

$$\mathbf{h}_t^e = \mathrm{ReLU}\left(\sum_{l=1}^{L} g_t^l W_l^e \mathbf{h}_t^l\right)$$

Handwriting Sequence Generation

- Data : (x_t, y_t, p_t)

Visualization by segments using
the ground truth of pen-tip location

Visualization by segments using
the states of z^2