Multi cell LSTM

Povilas Daniušis *

February 9, 2017

1 Introduction

In this document we formulate new extension of long short-term memory (LSTM) recurrent neural network.

1.1 Conventional LSTM

Let $x_t \in \mathbb{R}^{D_x}$ be input vectors. LSTM state $(c_t, h_t)^T$ is defined by:

$$\begin{aligned} i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \\ f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \\ o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o) \\ \tilde{c}_t &= \tanh(W^c x_t + U^c h_{t-1} + b^c)' \\ c_t &= f_t \bullet c_{t-1} + i_t \bullet \tilde{c}_t \\ h_t &= o_t \bullet \tanh(c_t), \end{aligned} \tag{1}$$

Where i_t, f_t, o_t are calles input, forget and output gates, $\tilde{c}_t$ - candidate cell vector, c_t - cell vector, and h_t - hidden state vector. All aforementioned variables are D_h- dimensional. By $\bullet$ we denote element-wise multiplication. Parameter count of LSTM is:

$$N = 4 \cdot D_h \cdot (D_x + D_h + 1), \tag{2}$$

and state is defined by $2D_h$ variables. c_t may be interpreted as internal memory of LSTM, while h_t - represent its content, exposed by the output gate.

*povilas.daniusis@gmail.com

1.2 Multi cell LSTM

The internal memory of LSTM, c_t, shares the same dimension with h_t. However, according to our intuition c_t should be able to store more information than h_t. We hypothesise, that extension of c_t may increase effectiveness of LSTM approach. We suggest multi-cell variant of LSTM (MCLSTM) with D_p cells, assembled into $D_h \times D_p$ matrix C_t. Internal attention p_t controls importance weights of individual cells (columns of C_t).

$$\begin{aligned}
i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \in \mathbb{R}^{D_h} \\
f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \in \mathbb{R}^{D_h} \\
o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o) \in \mathbb{R}^{D_h} \\
p_t &= \mathrm{softmax}(W^p x_t + U^p h_{t-1} + b^p) \in \mathbb{R}^{D_p} \\
\tilde{C}_t &= \tanh(W^c x_t + U^c h_{t-1} + b^c) 1^T \in \mathbb{R}^{D_h \times D_p} \\
C_t &= (f_t p_t^T) \bullet C_{t-1} + (i_t p_t^T) \bullet \tilde{C}_t \in \mathbb{R}^{D_h \times D_p} \\
h_t &= \frac{1}{D_p} o_t \bullet (\tanh(C_t) 1) \in \mathbb{R}^{D_h}
\end{aligned}, \tag{3}$$

Parameter count of multi cell LSTM is:

$$N = (4 \cdot D_h + D_p) \cdot (D_x + D_h + 1), \tag{4}$$

and state is defined by $D_h(D_p + 1)$ variables.

The implementation of MCLSTM can be downloaded from: `https://github.com/povidanius/multi_cell_lstm`

1.3 Kronecker LSTM

$$\begin{aligned}
(I_t, F_t, O_t) &= \sigma(X_t \otimes W^{i,f,o} + H_{t-1} \otimes U^{i,f,o} + B^{i,f,o}) \\
\tilde{C}_t &= \tanh(X_t \otimes W^c + H_{t-1} \otimes U^c + B^c) \\
C_t &= F_t \bullet C_{t-1} + I_t \bullet \tilde{C}_t \\
H_t &= P(O_t \bullet \tanh(C_t))Q,
\end{aligned} . \tag{5}$$

where $X_t : D_1^X \times D_2^X$, $W : D_1^W \times D_2^W$,$H_t : D_1^H \times D_2^H$,

$$IFOC : D_1^X D_1^W \times D_2^X D_2^W = D_1^H D_1^U \times D_2^H D_2^U$$

.