Multi cell LSTM

Povilas Daniušis *

February 3, 2017

1 Introduction

In this document we formulate new modification of long short-term memory.

1.1 Conventional LSTM

$$\begin{aligned} i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \\ f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \\ o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o) \\ \tilde{c}_t &= \tanh(W^c x_t + U^c h_{t-1} + b^c) \\ c_t &= f_t \bullet c_{t-1} + i_t \bullet \tilde{c}_t \\ h_t &= o_t \bullet \tanh(c_t) \end{aligned}, \tag{1}$$

1.2 Multi cell LSTM

We suggest variant of LSTM 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).

$$\begin{aligned} i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \in \mathbb{R}^{D_h} \\ f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \in \mathbb{R}^{D_h} \\ o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o) \in \mathbb{R}^{D_h} \\ p_t &= \text{softmax}(W^p x_t + U^p h_{t-1} + b^p) \in \mathbb{R}^{D_p} \\ \tilde{C}_t &= \tanh(W^c x_t + U^c h_{t-1} + b^c) 1^T \in \mathbb{R}^{D_h \times D_p} \\ C_t &= (f_t p_t^T) \bullet C_{t-1} + (i_t p_t^T) \bullet \tilde{C}_t \in \mathbb{R}^{D_h \times D_p} \\ h_t &= \frac{1}{D_p} o_t \bullet (\tanh(C_t) 1) \in \mathbb{R}^{D_h} \end{aligned}, \tag{2}$$

*povilas.daniusis@gmail.com