Wide and Deeper, Cheaper and Faster: Tensorized LSTMs for Sequence Learning

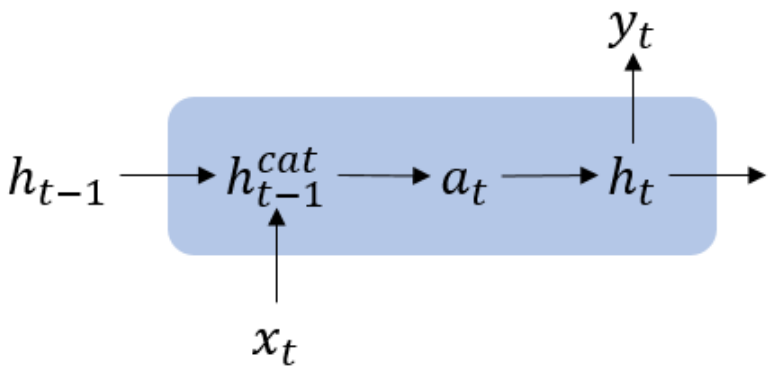
Zhen He, Shaobing Gao, Liang Xiao, Daxue Liu, Hangen He, and David Barber

Kuhwan Jeong [1]

[1] Department of Statistics, Seoul National University

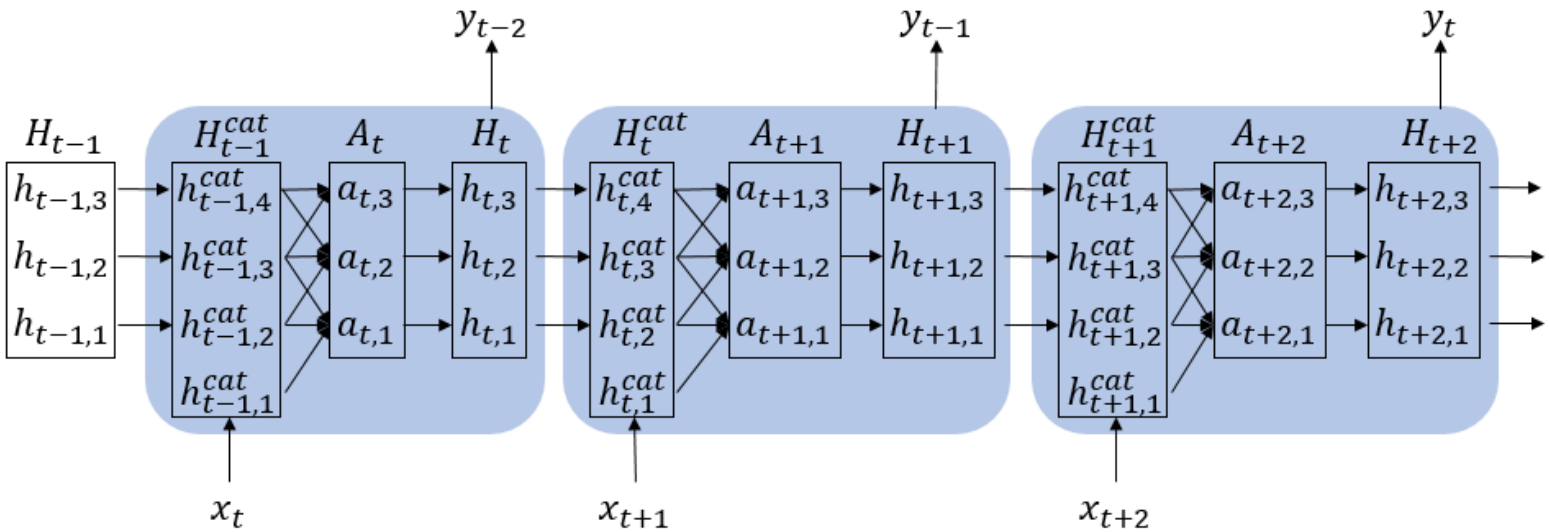
February, 2018

RNN



$$
\begin{aligned}
h_{t-1}^{cat} &= [x_t, h_{t-1}] \\
a_t &= W^h h_{t-1}^{cat} + b^h \\
h_t &= \phi(a_t) \\
y_t &= \varphi(W^y h_t + b^y)
\end{aligned}
$$

tRNN



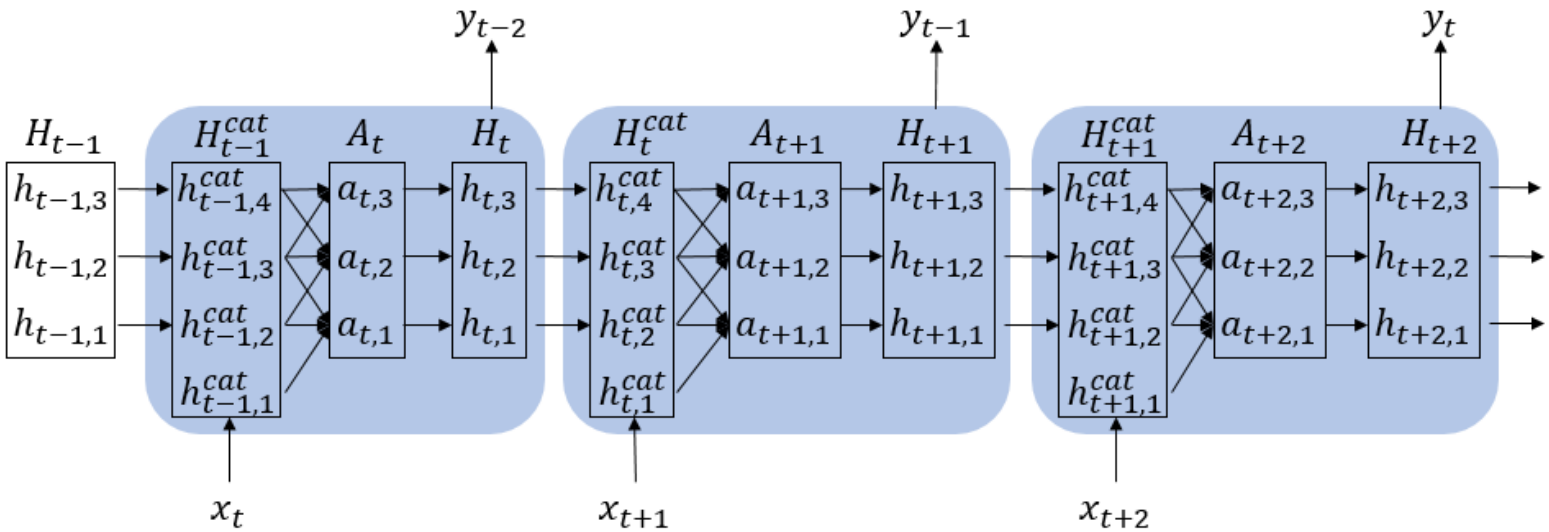
$$h_{t-1,p}^{cat} \propto \begin{cases} W^x x_t + b^x & \text{if } p = 1 \\ h_{t-1,p-1} & \text{otherwise} \end{cases}$$

$$A_t = H_{t-1}^{cat} \circledast \{W^h, b^h\}$$

$$G_t = \phi(A_t)$$

$$y_t = \varphi(W^y h_{t+L-1,P} + b^y)$$

tRNN



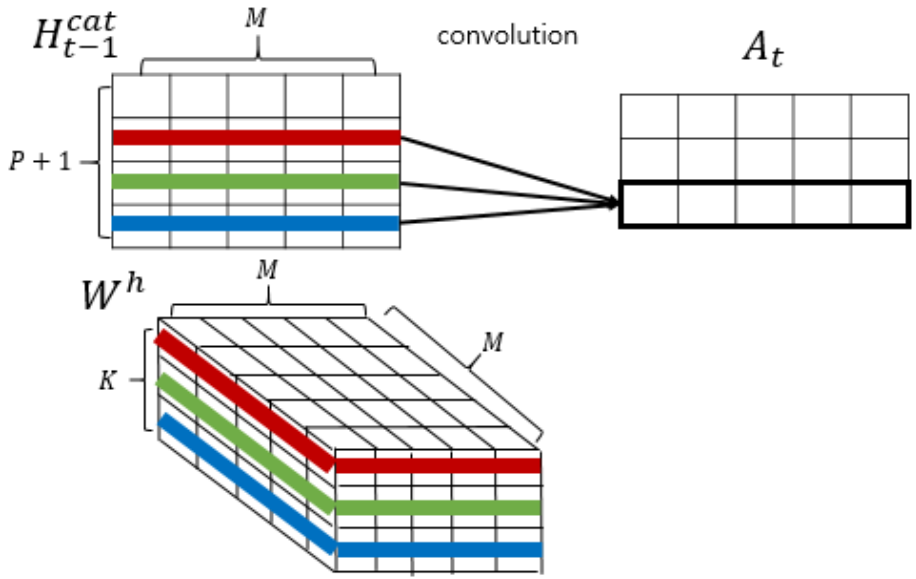
$$h_{t-1,p}^{cat} \propto \begin{cases} W^x x_t + b^x & \text{if } p = 1 \\ h_{t-1,p-1} & \text{otherwise} \end{cases}$$

$$A_t = H_{t-1}^{cat} \circledast \{W^h, b^h\}$$

$$G_t = \phi(A_t)$$

$$y_t = \varphi(W^y h_{t+L-1,P} + b^y)$$

tRNN



$$a_{t,p,m} = \sum_{k=1}^{K} \left(\sum_{i=1}^{M} W_{k,m,i} h_{t-1,p-\frac{K-1}{2}+k,i} \right) + b_m^h$$

- The tRNN can be widened by increasing the tensor size P, whilst the parameter number remains fixed.
- Unlike the sRNN of runtime complexity $O(TL)$, tRNN breaks down the runtime complexity to $O(T + L)$.

tLSTM

$$
\begin{aligned}
[A_t^g, A_t^i, A_t^f, A_t^o] &= H_{t-1}^{cat} \circledast \{W^h, b^h\} \\
[G_t, I_t, F_t, O_t] &= [\phi(A_t^g), \sigma(A_t^i), \sigma(A_t^f), \sigma(A_t^o)] \\
C_t &= G_t \circledcirc I_t + C_{t-1} \circledcirc F_t \\
H_t &= \phi(C_t) \circledcirc O_t
\end{aligned}
$$