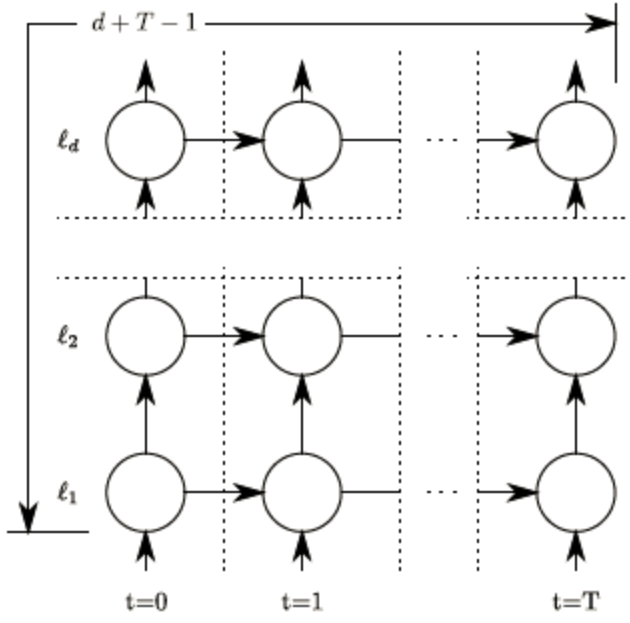
Recurrent Highway Network

Julian Georg Zilly, Rupesh Kumar Srivastava, Jan KoutnÃk, JÃ¼rgen Schmidhuber

Motivation

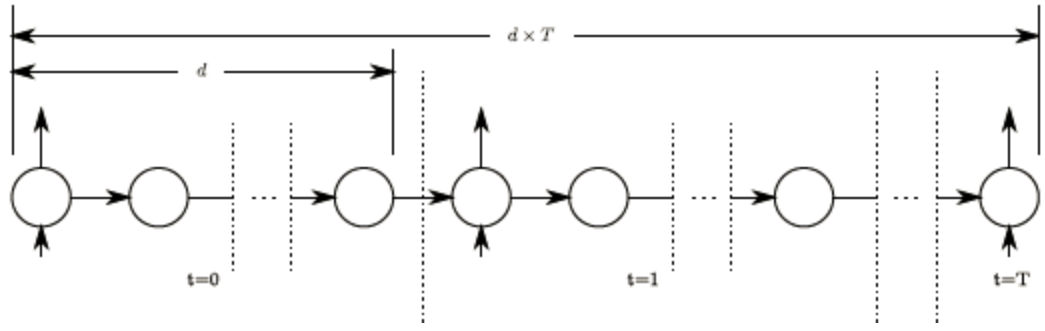
Conducting a extension to each LSTM state in space (horizontal) to make the gradient information flowing more efficient.

RNN



The longest credit assignment path is d+T-1. d is the stack number, and T is the time number.

Deep Transition RNN



The longest credit assignment path is d*T.

Training is more difficult.

Recurrent Highway Network (RHN)

Incorporting a highway network mechanism.

Doing highway multi-transition in each hidden state unit.

$$s_l^{[t]} = h_l^{[t]} \cdot t_l^{[t]} + s_{l-1}^{[t]} \cdot c_l^{[t]}$$

where

$$\begin{aligned}
\mathbf{h}_\ell^{[t]} &= tanh(\mathbf{W}_H \mathbf{x}^{[t]} \mathbb{I}_{\{\ell=1\}} + \mathbf{R}_{H_\ell} \mathbf{s}_{\ell-1}^{[t]} + \mathbf{b}_{H_\ell}), \\
\mathbf{t}_\ell^{[t]} &= \quad \sigma(\mathbf{W}_T \mathbf{x}^{[t]} \mathbb{I}_{\{\ell=1\}} + \mathbf{R}_{T_\ell} \mathbf{s}_{\ell-1}^{[t]} + \mathbf{b}_{T_\ell}), \\
\mathbf{c}_\ell^{[t]} &= \quad \sigma(\mathbf{W}_C \mathbf{x}^{[t]} \mathbb{I}_{\{\ell=1\}} + \mathbf{R}_{C_\ell} \mathbf{s}_{\ell-1}^{[t]} + \mathbf{b}_{C_\ell}),
\end{aligned}$$

l is the number of highway layers.

Conclusion

RHN can get better performance in some tasks by making less parameters. (fewer hidden dim)