MEMORY VISUALIZATION FOR GATED RECURRENT NEURAL NETWORKS IN SPEECH RECOGNITION

Zhiyuan Tang, Ying Shi, Dong Wang, Yang Feng, Shiyue Zhang

Center for Speech and Language Technologies (CSLT), Tsinghua University

Abstract

-- Visualizing the behavior of LSTM and GRU when performing speech recognition tasks:

1. Activation patterns
2. Temporal trace
3. Memory robustness

-- Modifications inspired by the visualization:

1. lazy cell update in LSTM
2. shortcut connections for residual learning

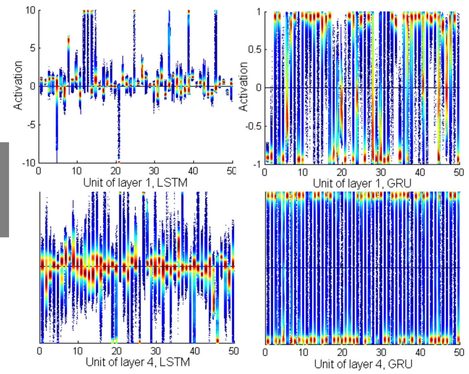
LSTM & GRU

$$
\begin{aligned}
i_t &= \sigma(W_{ix}x_t + W_{im}m_{t-1} + V_{ic}c_{t-1}) \\
f_t &= \sigma(W_{fx}x_t + W_{fm}m_{t-1} + V_{fc}c_{t-1}) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1}) \\
o_t &= \sigma(W_{ox}x_t + W_{om}m_{t-1} + V_{oc}c_t) \\
m_t &= o_t \odot h(c_t).
\end{aligned}
$$

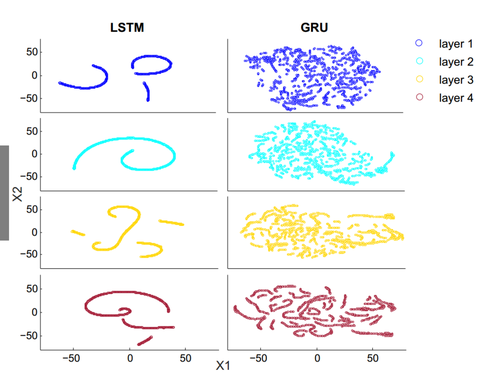
$$
\begin{aligned}
i_t &= \sigma(W_{ix}x_t + W_{ic}c_{t-1}) \\
f_t &= 1 - i_t \\
o_t &= \sigma(W_{ox}x_t + W_{oc}c_{t-1}) \\
m_t &= o_t \odot c_{t-1} \\
c_t &= f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_t).
\end{aligned}
$$

Visualization

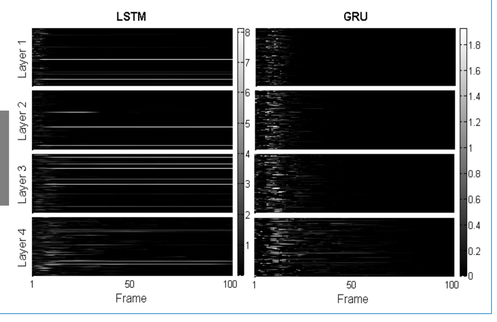
Activation patterns



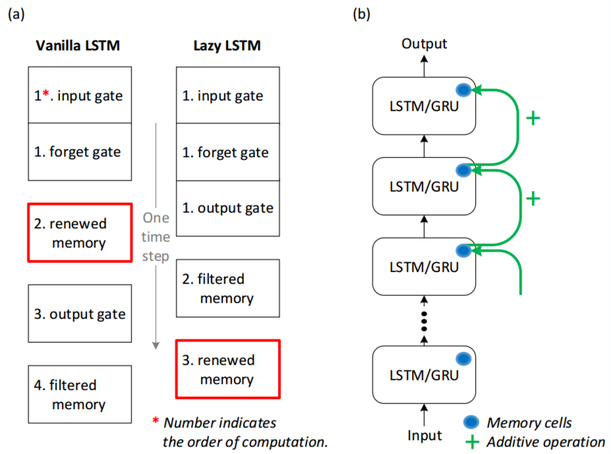
Temporal trace



Memory robustness



Modifications



(a) Lazy cell update: GRU updates cells at the final step, while LSTM updates cells before computing output gates.

(b) Shortcut connection for residual learning: gates at high-level layers show a similar pattern, cells in high-level layers are mostly learned by residual.

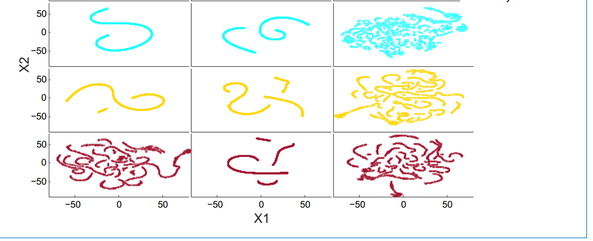
Experiment set

Date:	WSJ database
GMMs:	MFCC, 3377 pdfs
RNNs:	Fbank
	LSTM/GRU, 512 cells
	1/2/4/6 layers

Experiment Results

	WER%	
Recurrent Layers	Baseline	Lazy Update
1	10.96	10.18
2	9.97	9.48
4	9.67	9.10

		WER%	
System	Recurrent Layers	Baseline	Residual Learning
LSTM	4	9.67	9.53
	6	9.47	9.33
GRU	4	9.32	9.23
	6	9.32	9.10



Conclusions

-- LSTM and GRU use different ways to encode information and the information in GRU is more distributed. LSTM possesses a long-term memory but it is also noise-sensitive.

-- Inspired by these observations, we introduced two modifications to enhance gated RNNs: lazy cell update and short connections for residual learning, and both provide interesting performance improvement.

ICASSP, New Orleans, USA. Mar 2017.