What Does It Mean to Understand?

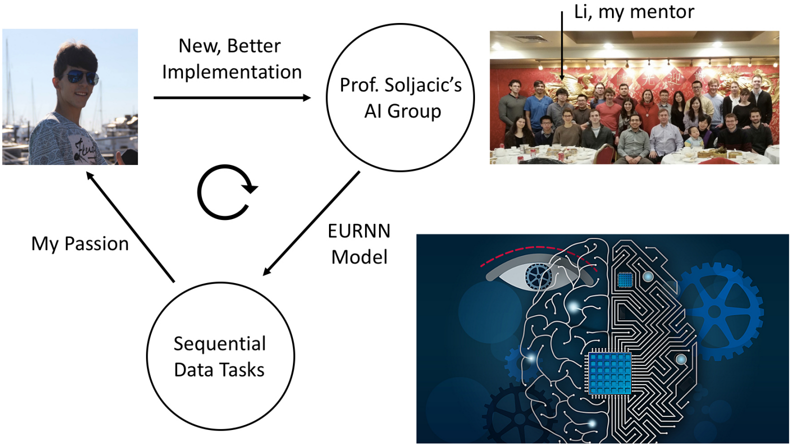
Improving the Performance of Unitary Recurrent Neural Networks and Applying Them to the Automatic Text Understanding Problem

Ivan Ivanov

Research Science Institute

Under the Direction of Li Jing
Massachusetts Institute of Technology

The Purpose Behind **Artificial** Intelligence

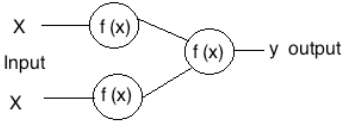


(MIT News) and (Prof. Soljacic's group)

The Artificial **Neural** Network

- An Analogy

Biological Neural Network

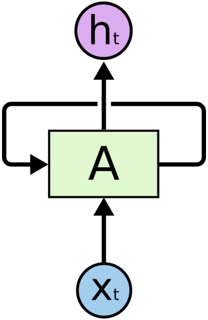


Artificial Neural Network

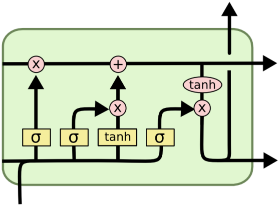
(Dasan)

The Artificial **Recurrent** Neural Network

- Concept of Recurrence



General Model

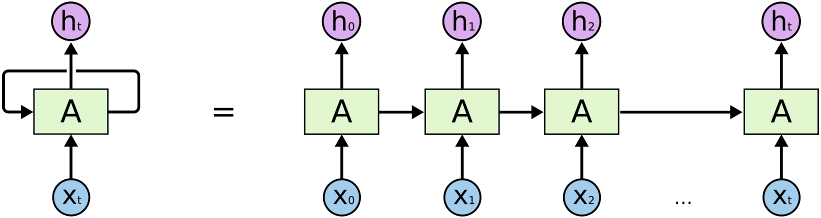


Long Short-Term Memory (LSTM) Model

(Olah)

The Artificial **Recurrent** Neural Network

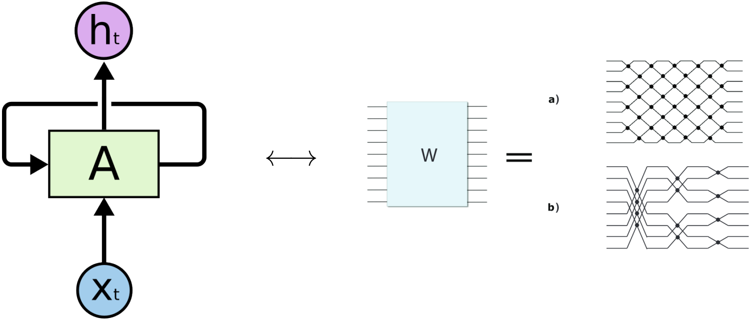
- Concept of Recurrence



(Olah)

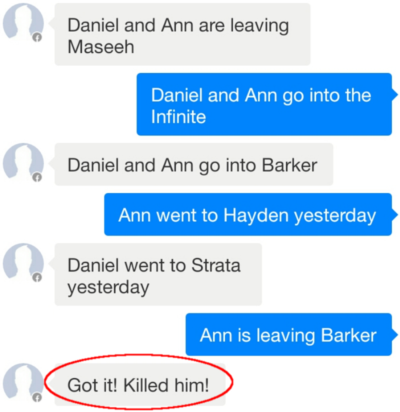
The Artificial **Unitary** Recurrent Neural Network

- The Unitary Matrix



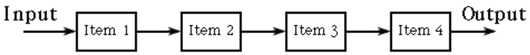
(Jing et. al)

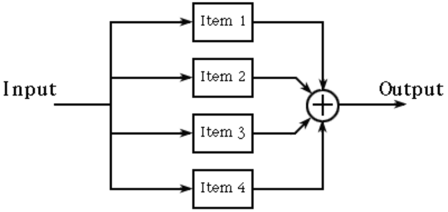
The Artificial **Unitary** Recurrent Neural Network



- Dr. Sillman: *Where is Daniel?*

Optimizations of the model

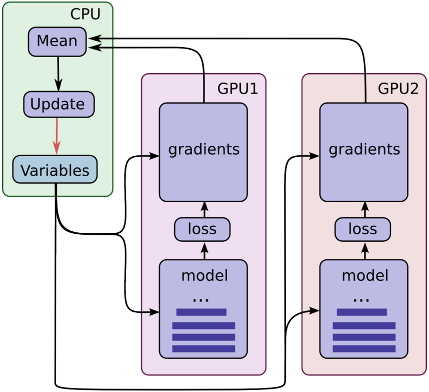




(Volans)

- Parallelization

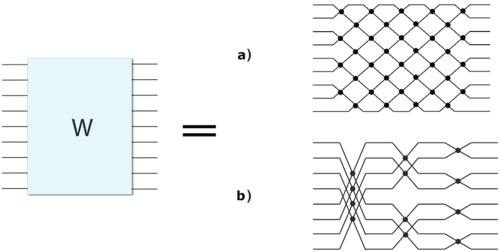
Optimizations of the model



(PipelineAI)

- Parallelization
- TensorFlow Adaptation

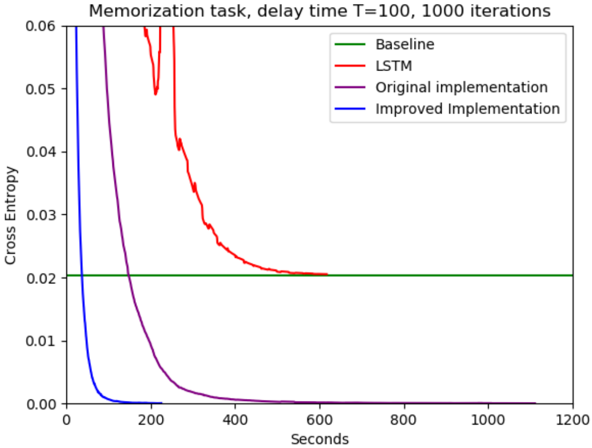
Optimizations of the model



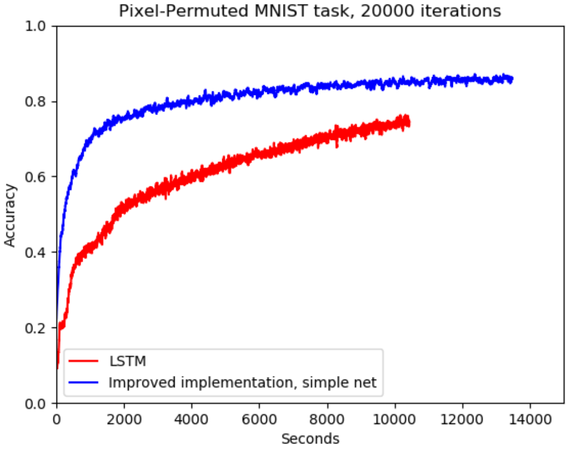
(Jing et al.)

- Parallelization
- TensorFlow Adaptation
- Hyperparameter Expansion

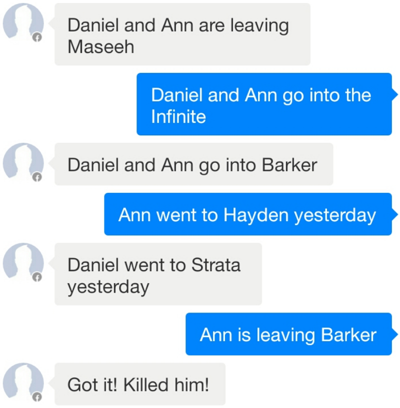
Benchmark: Memorization Task



Benchmark: Handwritten Digit Recognition Task

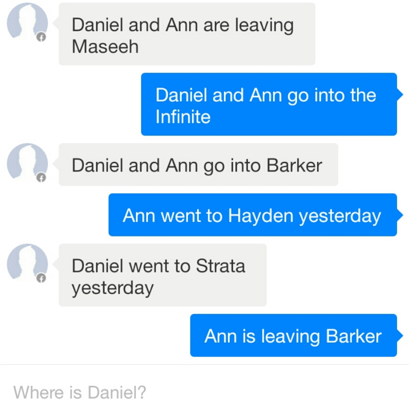


Automatic Text Understanding

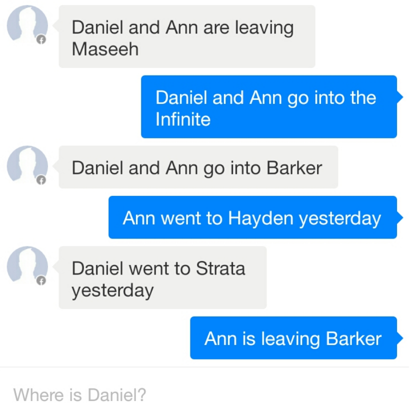


- Dr. Sillman: *Where is Daniel?*

Automatic Text Understanding

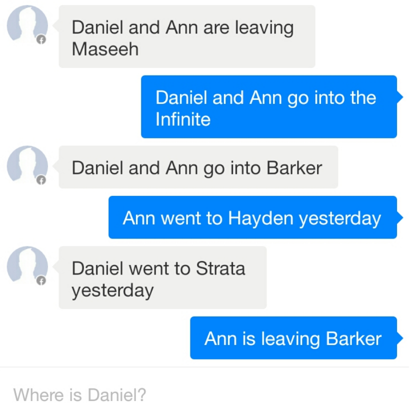


Automatic Text Understanding



Daniel	and	...	is	Where
0	1	...	16	17

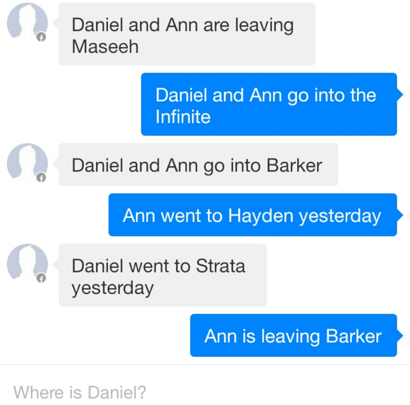
Automatic Text Understanding



Daniel	and	...	is	Where
0	1	...	16	17

Where is Daniel? $\longrightarrow$ *(17, 16, 0)*

Automatic Text Understanding



Daniel	and	...	is	Where
0	1	...	16	17

$$\textit{Where is Daniel?} \longrightarrow (17,\ 16,\ 0)$$

$$\downarrow$$

$$([0,0,\ldots,0,1],\ [0,0,\ldots,1,0],\ [1,0,\ldots,0,0])$$

Automatic Text Understanding

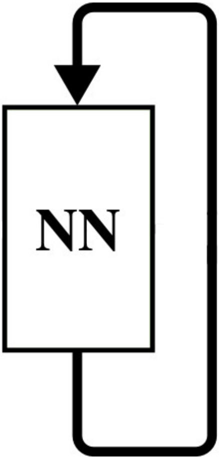
$$\begin{array}{l} 0 \rightarrow \\ 1 \rightarrow \\ 0 \rightarrow \\ \vdots \\ 0 \rightarrow \\ 0 \rightarrow \\ 0 \rightarrow \end{array}$$

Automatic Text Understanding

$$\begin{array}{l} 0 \rightarrow \\ 1 \rightarrow \\ 0 \rightarrow \\ \vdots \\ 0 \rightarrow \\ 0 \rightarrow \\ 0 \rightarrow \end{array}$$

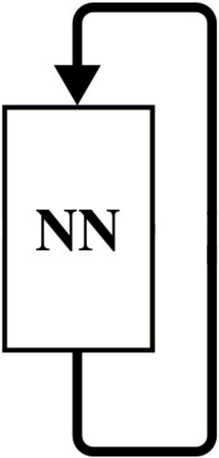
Automatic Text Understanding

$0 \rightarrow$
$1 \rightarrow$
$0 \rightarrow$
$\vdots$
$0 \rightarrow$
$0 \rightarrow$
$0 \rightarrow$



Automatic Text Understanding

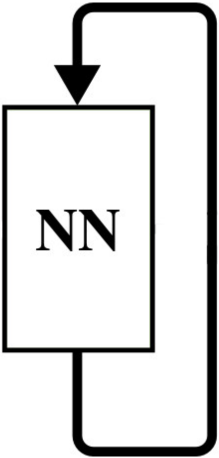
$0 \rightarrow$
$1 \rightarrow$
$0 \rightarrow$
$\vdots$
$0 \rightarrow$
$0 \rightarrow$
$0 \rightarrow$



$\rightarrow$ 0.12 (12%)
$\rightarrow$ 0.01 (1%)
$\rightarrow$ 0.07 (7%)
$\vdots$
$\rightarrow$ 0.24 (24%)
$\rightarrow$ 0.56 (56%)
$\rightarrow$ 0.00 (0%)

Automatic Text Understanding

$0 \rightarrow$		$\rightarrow$ 0.12 (12%)	Daniel
$1 \rightarrow$		$\rightarrow$ 0.01 (1%)	and
$0 \rightarrow$		$\rightarrow$ 0.07 (7%)	Ann
$\vdots$		$\vdots$	$\vdots$
$0 \rightarrow$		$\rightarrow$ 0.24 (24%)	Strata
$0 \rightarrow$		$\rightarrow$ 0.56 (56%)	**Barker**
$0 \rightarrow$		$\rightarrow$ 0.00 (0%)	Where



bAbI Tasks Dataset

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

(Facebook)

bAbI Tasks Dataset

Task	Our model	LSTM
1 - Single Supporting Fact	50.5%	**52.0%**
2 - Two Supporting Facts	**31.8%**	15.1%
3 - Three Supporting Facts	**25.4%**	19.1%
4 - Two Arg. Relations	71.2%	**73.5%**
5 - Three Arg. Relations	**67.1%**	34.4%
6 - Yes/No Questions	**52.9%**	50.5%
7 - Counting	**71.3%**	56.5%
8 - Lists/Sets	**68.2%**	38.8%
9 - Simple Negation	61.8%	**63.8%**
10 - Indefinite Knowledge	**46.0%**	45.1%
11 - Basic Coreference	72.3%	**74.1%**
12 - Conjunction	73.4%	**76.1%**
13 - Compound Coreference	**94.0%**	83.0%
14 - Time Reasoning	**36.4%**	18.6%
15 - Basic Deduction	**55.0%**	21.2%
16 - Basic Induction	**48.8%**	32.2%
17 - Positional Reasoning	48.4%	**50.6%**
18 - Size Reasoning	**89.5%**	89.2%
19 - Path Finding	**7.9%**	6.6%
20 - Agents Motivations	**95.5%**	90.6%
Mean Performance	**58.4%**	49.6%

bAbI Tasks Dataset

Task	Our model	LSTM
1 - Single Supporting Fact	50.5%	**52.0%**
2 - Two Supporting Facts	**31.8%**	15.1%
3 - Three Supporting Facts	**25.4%**	19.1%
4 - Two Arg. Relations	71.2%	**73.5%**
5 - Three Arg. Relations	**67.1%**	34.4%
6 - Yes/No Questions	**52.9%**	50.5%
7 - Counting	**71.3%**	56.5%
8 - Lists/Sets	**68.2%**	38.8%
9 - Simple Negation	61.8%	**63.8%**
10 - Indefinite Knowledge	**46.0%**	45.1%

Task	Our model	LSTM
11 - Basic Coreference	72.3%	**74.1%**
12 - Conjunction	73.4%	**76.1%**
13 - Compound Coreference	**94.0%**	83.0%
14 - Time Reasoning	**36.4%**	18.6%
15 - Basic Deduction	**55.0%**	21.2%
16 - Basic Induction	**48.8%**	32.2%
17 - Positional Reasoning	48.4%	**50.6%**
18 - Size Reasoning	**89.5%**	89.2%
19 - Path Finding	**7.9%**	6.6%
20 - Agents Motivations	**95.5%**	90.6%
Mean Performance	**58.4%**	49.6%

bAbI Tasks Dataset

Task	Our model	LSTM	Task	Our model	LSTM
1 - Single Supporting Fact	50.5%	**52.0%**	11 - Basic Coreference	72.3%	**74.1%**
2 - Two Supporting Facts	**31.8%**	15.1%	**12 - Conjunction**	73.4%	**76.1%**
3 - Three Supporting Facts	**25.4%**	19.1%	13 - Compound Coreference	**94.0%**	83.0%
4 - Two Arg. Relations	71.2%	**73.5%**	14 - Time Reasoning	**36.4%**	18.6%
5 - Three Arg. Relations	**67.1%**	34.4%	15 - Basic Deduction	**55.0%**	21.2%
6 - Yes/No Questions	**52.9%**	50.5%	16 - Basic Induction	**48.8%**	32.2%
7 - Counting	**71.3%**	56.5%	17 - Positional Reasoning	48.4%	**50.6%**
8 - Lists/Sets	**68.2%**	38.8%	18 - Size Reasoning	**89.5%**	89.2%
9 - Simple Negation	61.8%	**63.8%**	19 - Path Finding	**7.9%**	6.6%
10 - Indefinite Knowledge	**46.0%**	45.1%	20 - Agents Motivations	**95.5%**	90.6%
			Mean Performance	**58.4%**	49.6%

Conclusion

- Five times the efficiency of the original implementation
- Greater accuracy than state-of-the-art model on bAbI tasks dataset
- Introduction of the theoretical model to a real-life task

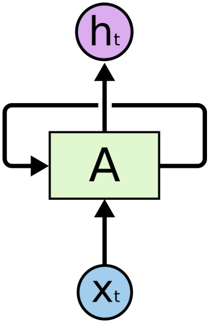
Future work

- Decomposition model improvements
- Low-level operations optimization
- Application for speech recognition

Acknowledgements

- Li Jing
- Rumen Dangovski
- Dr. Jenny Sendova
- Andrew Jin, Charles Tam, Stanislav Atanasov, William McInroy, Hristo Stoyanov, Milen Ferev
- RSI, CEE, MIT
- America for Bulgaria Foundation, International Foundation "Sts. Cyril & Methodius"

The Recurrent Neural Network



$$m^{(t)} = \sigma(U * x^{(t)} + W * m^{(t-1)})$$

$$h^{(t)} = W * m^{(t)} + b$$

Exploding and Vanishing Gradients Problems

Training rule: $W_{i,j} - \lambda * \frac{\partial C}{\partial W_{i,j}}$

$$\frac{\partial C}{\partial h^{(t)}} = \frac{\partial C}{\partial h^{(T)}} \frac{\partial h^{(T)}}{\partial h^{(t)}} = \frac{\partial C}{\partial h^{(T)}} \prod_{k=t}^{T-1} \frac{\partial h^{(k+1)}}{\partial h^{(k)}} = \frac{\partial C}{\partial h^{(T)}} \prod_{k=t}^{T-1} D^{(t)} W$$

Jing et al.'s Approach

- General representation

$$\begin{aligned} W_n &= DR_{2,1}^{-1} R_{3,1}^{-1} \dots R_{N,N-2}^{-1} R_{N,N-1}^{-1} \\ &= DR_{2,1}^{'} R_{3,1}^{'} \dots R_{N,N-2}^{'} R_{N,N-1}^{'} \end{aligned}$$

Jing et al.'s Approach

- Simple Net Decomposition Model

$$
\begin{aligned}
W &= D(R^{(1)}_{1,2} R^{(1)}_{3,4} \ldots R^{(1)}_{N/2-1,N/2}) \times \\
&\times (R^{(2)}_{2,3} R^{(2)}_{4,5} \ldots R^{(2)}_{N/2-1,N/2-1}) \times \ldots \\
&= D F^{(1)}_a F^{(2)}_b \ldots F^{(L)}_b
\end{aligned}
$$

$$
\begin{aligned}
F^{(l)}_a &= R^{(l)}_{1,2} R^{(l)}_{3,4} \ldots R^{(l)}_{N/2-1,N/2} \\
F^{(l)}_b &= R^{(l)}_{2,3} R^{(l)}_{4,5} \ldots R^{(l)}_{N/2-1,N/2-1}
\end{aligned}
$$

Jing et al.'s Approach

- Lightweight Decomposition Model

$$W = DF_1 F_2 \dots F_{log(N)}$$

F_i - rotation matrices for $(2kp + j, (2k+1)p + j)$,
$p = N/2t, \; k \in 0, \dots, 2^{i-1}$, and $j \in 1, \dots, p$

Jing et al.'s Approach

$$\mathbf{F}x = v_1 * x + v_2 * permute(x)$$

- Simple Net Decomposition Model

$$v_1 = (e^{i\phi_1^{(l)}} \cos\theta_1^{(l)},\ \cos\theta_1^{(l)},\ e^{i\phi_2^{(l)}} \cos\theta_2^{(l)},\ \cos\theta_2^{(l)}, \ldots)$$
$$v_2 = (-e^{i\phi_1^{(l)}} \sin\theta_1^{(l)},\ \sin\theta_1^{(l)},\ -e^{i\phi_2^{(l)}} \sin\theta_2^{(l)},\ \sin\theta_2^{(l)}, \ldots)$$
$$permute(x) = (x_2, x_1, x_4, x_3, x_6, x_5, \ldots)$$

- Lightweight Decomposition Model

$$v_1 = (e^{i\phi_1^{(l)}} \cos\theta_1^{(l)},\ e^{i\phi_2^{(l)}} \cos\theta_2^{(l)}, \ldots,\ \cos\theta_1^{(l)},\ \cos\theta_2^{(l)}, \ldots)$$
$$v_2 = (-e^{i\phi_1^{(l)}} \sin\theta_1^{(l)},\ -e^{i\phi_2^{(l)}} \sin\theta_2^{(l)}, \ldots,\ \sin\theta_1^{(l)},\ \sin\theta_2^{(l)}, \ldots)$$
$$permute(x) = (x_{N/2} + 1, x_{N/2} + 2, \ldots, x_N, x_1, x_2, \ldots)$$