Exploiting Randomness in Neural Networks

Daniele Di Sarli

Mauriana Pesaresi seminars - 2020
Recurrent Neural Network
Backpropagation Through Time

\[ \text{error} = (\text{output} - \text{expected})^2 \\]
PREDICTION

PATTERNS, INTERACTIONS

..., 3, 2, 1.5, 0.75, 1, -2.3, 4, ...
Reservoir

Readout
Echo State Network
State Space $\mathbb{R}^{NR}$

(a) 8, 9, 5, 3, 5, 4, 8, 3, 1, 9, 5
(b) 7, 7, 3, 4, 4, 3, 1, 2, 1, 9, 5
(c) 4, 8, 1, 5, 6, 0, 7, 8, 2, 2, 3
(d) 2, 4, 3, 2, 0, 2, 0, 4, 2, 2, 3
Cover’s theorem

State Space $\mathbb{R}^{N_R}$
Echo State Property

\[ \forall s_N(u) = [u(1), \ldots, u(N)] \in (\mathbb{R}^{N_U})^N, \]
\[ \forall x, x' \in \mathbb{R}^{N_R} : \]
\[ \| \hat{\tau}(s_N(u), x) - \hat{\tau}(s_N(u), x') \| \to 0 \text{ as } N \to \infty \]
Echo State Network starter pack

1. **Randomly initialize** the weights (sparse)

2. Rescale the weights to guarantee **contractivity** of the state transition function (=> ESP)

3. Feed data, collect states

4. Compute **optimal** linear regression parameters

\[
W_{out} = \bar{Y}X^T(XX^T + \lambda I)^{-1}
\]
«RC [...] provides explanations of why biological brains can carry out accurate computations with an “inaccurate” and noisy physical substrate»

— Lukoševičius et al.

In the primary visual cortex, «computations are performed by complex dynamical systems while information about results of these computations is read out by simple linear classifiers.»

— Nikolić et al.
My work
Natural Language Processing

<table>
<thead>
<tr>
<th>Model</th>
<th>CO₂ emissions (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td></td>
</tr>
<tr>
<td>GRU</td>
<td></td>
</tr>
<tr>
<td>Transformer</td>
<td></td>
</tr>
<tr>
<td>Transformer w/ neural arch. search</td>
<td>600000</td>
</tr>
</tbody>
</table>

Car, avg incl. fuel, 1 lifetime

Text Classification pipeline

input sequence
(word embeddings)  →  RNN  →  sentence embedding  →  linear classifier

Training
Text Classification pipeline

\[ 'My \text{ input sentence}' \rightarrow \text{ESN} \rightarrow \text{sentence embedding} \rightarrow \text{linear classifier} \]
Question Classification

What was the name of the first Russian astronaut to do a spacewalk?

HUMAN

What's the tallest building in New York City?

LOCATION

... also ABBREVIATION, ENTITY, DESCRIPTION, and NUMERIC VALUE
Improvements are needed

- **Bidirectional**
- Attention
- Multi-ring

*What's the tallest building in New York City?*
Improvements are needed

• Bidirectional
• **Attention**
• Multi-ring

What's the tallest building in New York City?
Improvements are needed

• Bidirectional
• Attention
• Multi-ring
Results

Accuracy

Ada-CNN  Bi-LSTM  Paragraph Vector  Transformer + CNN  Bi-GRU  Bi-ESN  Bi-ESN (ensemble)  Bi-ESN-Att

200M+ params, heavy transfer learning

ours  < 1.6M params
Results

Accuracy

Bi-GRU
Bi-ESN
Bi-ESN (ensemble)
Bi-ESN-Att

200M+ params, heavy transfer learning
ours
< 1.6M params

Training time

Bi-GRU
Bi-ESN
Bi-ESN (ensemble)
Bi-ESN-Att

7.5 min
6 sec

Ada-CNN
Bi-LSTM
Paragraph Vector
Transformer + CNN
Bi-GRU
Bi-ESN
Bi-ESN (ensemble)
Bi-ESN-Att

ours
< 1.6M params

200M+ params, heavy transfer learning
How old was the youngest president of the United States?
When was Ulysses S. Grant born?
Who invented the instant Polaroid camera?
What is nepotism?
Where is the Mason/Dixon line?
What is the capital of Zimbabwe?
What are Canada's two territories?
What was the name of the plane Lindbergh flew solo across the Atlantic?
What is genocide?
What continent is Argentina on?
What monastery was raided by Vikings in the late eighth century?
What is an earthquake?
Where is the tallest roller coaster located?
What are enzymes?
Who discovered oxygen?
What is bangers and mash?
What is the name given to the Tiger at Louisiana State University?
Where are the British crown jewels kept?
Who was the first person to reach the North Pole?
What is an ulcer?
What is vertigo?
What is the spirometer test?
When is the official first day of summer?
What does the abbreviation SOS mean?
What is the smallest bird in Britain?
Who invented Trivial Pursuit?
What gases are in the troposphere?
Which country has the most water pollution?
What is the scientific name for elephant?
Who is the actress known for her role in the movie "Gypsy"?
What breed of hunting dog did the Beverly Hillbillies own?
The rainiest place on Earth?
Who was the first African American to win the Nobel Prize in literature?
When is St. Patrick's Day?
What was FDR's dog's name?
What colors need to be mixed to get the color pink?
What is the most popular sport in Japan?
What is the active ingredient in baking soda?
When was Thomas Jefferson born?
How cold should a refrigerator be?
When was the telephone invented?
Wrap up

• A path towards efficient, effective ML models must be taken

• Heavier understanding/exploitation of the architectural properties of RNN models can help towards that goal

• Analysis is preliminary, but WIP results are encouraging
References


