Neural networks with external memory


Daniel Majoral López
Outline

- Introduction
- Neural Turing machine
- Differentiable neural computer
- Conclusions
Introduction

Neural Networks
Introduction

Neural Networks
Introduction

Neural Networks

Input → Hidden → Output

$\sum \varphi$ → $O$

$w_0, w_1, w_2, w_3, \ldots, w_n$

$x_0 = 1, x_1, x_2, x_3, \ldots, x_n$
Training Neural Networks

- Training a neural network consists in finding a set of weights between nodes such that the output of the network is as close as possible to the desired output.
Introduction

**Backpropagation**

- The backpropagation algorithm is used for training a network, it assigns the amount of responsibility in the error to each node.
The backpropagation algorithm is used for training a network, which assigns amount of responsibility for the error to each node.
The backpropagation algorithm is used for training a network, which assigns amount of responsibility for the error to each node.
Gradient descend

- In conjunction with backpropagation an optimization algorithm is used to find the minimum of the error. Typically gradient descend:
Introduction

Turing Machine

- Infinite tape zeros and ones
- Head that reads and writes
- Finite state register
- Finite set of instructions
Neural Turing Machines

**Scheme NTM**

Diagram showing the components of a Neural Turing Machine (NTM) including external input/output, a controller, read heads, write heads, and memory.
NTM Heads

Two ways to read or write memory:

- Focus by content: The controller emits a key vector, which is compared to content in memory by a similarity measure. The similarity becomes a probability weight for each location.

- Focus by location: Rotational shift of weighting, if the focus is in one location the shift puts the focus in next location. I.e. [0 1 0 0] goes to [0 0 1 0]
Neural Turing Machines

Copy Task

Figure 8: NTM and LSTM Generalisation for the Repeat Copy Task. NTM generalises almost perfectly to longer sequences than seen during training. When the number of repeats is increased it is able to continue duplicating the input sequence fairly accurately; but it is unable to predict when the sequence will end, emitting the end marker after the end of every repetition beyond the eleventh. LSTM struggles with both increased length and number, rapidly diverging from the input sequence in both cases.
Differentiable neural computer

- Like a computer uses his memory to compute complex data structures, but learns to do it from data.

- Tasks already solved by symbolic artificial intelligence but out of reach for artificial neural networks.

- Improvements in memory management over Neural Turing Machine.
DNC architecture

- Controller deep neural network with LSTM.
- External memory with three distinct forms of differentiable attention.
- Distributions of weights to access memory
Three forms of attention

- Reading by Content Lookup: The controller emits a key vector, which is compared to the content by a similarity measure.

- Reading by Temporal Link Matrix: $L(i,j)$ close to 1 if $i$ was the next location written after $j$ and to 0 otherwise.

- Writing memory: “Usage” of each location represented by a number between 0 and 1.
Architecture scheme
Tasks performed

1. Answering elemental questions
2. Answering graph questions:
   a. Traversal
   b. Shortest path
   c. Inference
3. Solving puzzles
Traversal question
Methods graphs Tasks

- Curriculum learning with increasing complexity.

- Supervised learning

- For learning shortest path takes samples from the optimal policy !!!
Family tree question
Block puzzle experiments

**a** Weightings

**b** Goal T constraints

**c** Board states

**d** Planned action decodings

**e** t-SNE location goal labels
Methods puzzle task

- Curriculum learning with increasing complexity.
- Trained with Reinforcement learning.
- Reward function is the number of constraints satisfied minus penalization for non valid moves.
DNC solving a puzzle
Probability of optimal solution

<table>
<thead>
<tr>
<th>Minimum Required Moves</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Constraints</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>94</td>
<td>95</td>
<td>95</td>
<td>93</td>
<td>94</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>79</td>
<td>93</td>
<td>97</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>63</td>
<td>78</td>
<td>85</td>
<td>92</td>
<td>94</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>46</td>
<td>58</td>
<td>76</td>
<td>81</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
<td>33</td>
<td>46</td>
<td>62</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>22</td>
<td>32</td>
<td>51</td>
<td>65</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>34</td>
<td>17</td>
<td>18</td>
<td>30</td>
<td>44</td>
<td>50</td>
</tr>
</tbody>
</table>

a. DNC Percent Optimal

<table>
<thead>
<tr>
<th>Minimum Required Moves</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Constraints</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>48</td>
<td>47</td>
<td>48</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>38</td>
<td>34</td>
<td>34</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>42</td>
<td>43</td>
<td>46</td>
<td>44</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>22</td>
<td>18</td>
<td>14</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>10</td>
<td>3</td>
<td>0.47</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>4.7</td>
<td>1.1</td>
<td>0.16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>3</td>
<td>1.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

b. LSTM Percent Optimal
Comparing with Neural Turing Machine

- The NTM has no mechanism to ensure that blocks of allocated memory do not overlap and interfere.

- The NTM has no way of freeing locations that have already been written.

- The sequential information is lost when the NTM jumps to a not consecutive memory location (content addressing).
Conclusions

- An external memory can increase the capacity of neural networks. Performs tasks that a standard network with LSTM is not able to do.

- External memory will give multi-purpose capacity to neural networks but still not able to generalize learning.

- An idea will be try to implement better associative recall.
Thank You!