Learning Hierarchical Information Flow with Recurrent Neural Modules

Danijar Hafner¹, Alex Irpan¹, James Davidson¹, Nicolas Heess²
¹ Google Brain, ² DeepMind
NIPS 2017 #3374

1. Contribution

Learn connectivity rather than manually defining the layer structure for the task.
Brain-inspired sequence model consisting of communicating RNN modules.
Discovers skip-connections, feedback loops, and novel connectivity patterns.
Explored dynamic and static reading mechanisms.
Our model generalizes better and outperforms stacked GRUs on 3 sequential tasks.

2. Motivation

Neocortex often described as hierarchy but there are many side-connections and feedback loops:
Areas communicate both directly and indirectly via the thalamus. We focus on the latter here.
Modules communicating via a routing center include hierarchy as a special case.

3. Method: ThalNet

Multiple recurrent modules share their features via a routing center (concatenation of features):
Modules observe the previous center value using dynamic or static reading mechanisms:

4. Findings

ThalNet learns hierarchical information flow, skip-connections, and long feedback pathways:
Learned reading weights for ThalNet with 4 modules and different reading mechanisms.

5. Performance

<table>
<thead>
<tr>
<th>Task</th>
<th>GRU (params)</th>
<th>ThalNet (params)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permuted MNIST</td>
<td>65% (50k)</td>
<td>92% (50k)</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>50% (50k)</td>
<td>54% (50k)</td>
</tr>
<tr>
<td>Text8</td>
<td>1.40 BPC (14M)</td>
<td>1.40 BPC (12M)</td>
</tr>
</tbody>
</table>

6. Conclusion

Similar connectivity is learned for the same task.
Modularity and reading bottleneck regularize the model and improve generalization.
The training time is about 2-3x that of the baseline.
Other recurrent models might benefit from long feedback loops learned by ThalNet.
Provides framework for multi-task learning and online architecture search.

Contact: mail@danijar.com

Figure adapted from: Gross et al. 1993. Inferior temporal cortex as a pattern recognition device.