Learning to Transduce with Unbounded Memory

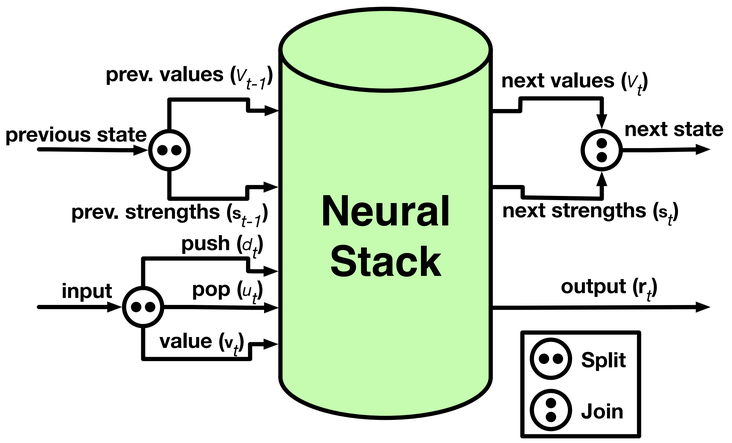
Edward Grefenstette **Karl Moritz Hermann** **Mustafa Suleyman** **Phil Blunsom**

Google DeepMind

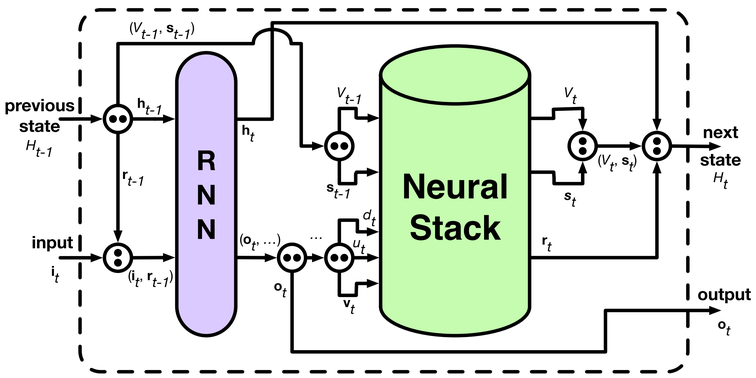
Three New Continuous Memories

Unbounded continuous memories simulating **stacks**, **queues**, and **double-ended queues** can be connected to recurrent networks of any topology to aid with transduction, record keeping, and other algorithmic tasks.

Neural Stack Schematics

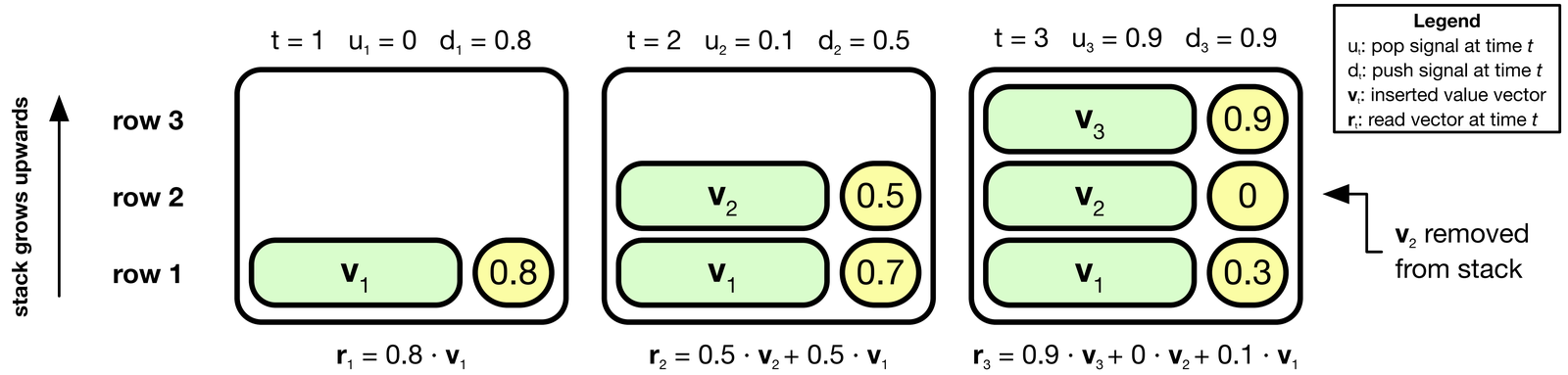


Neural Stack as a Recurrent Layer

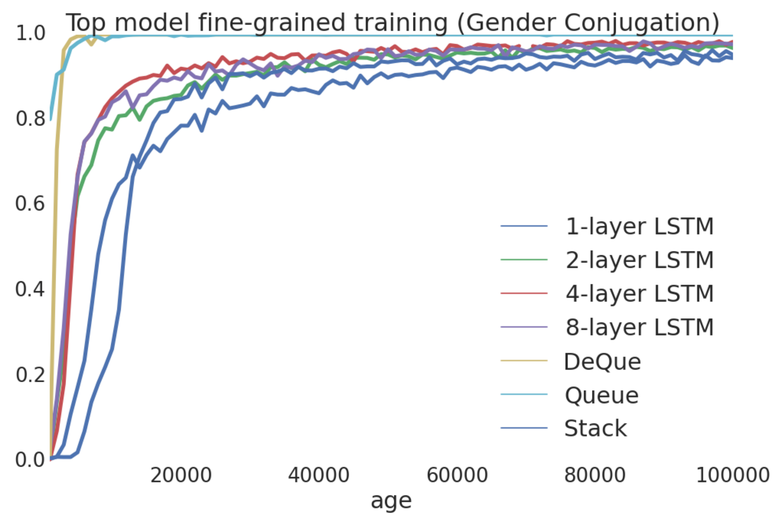


Controlling a Neural Memory

Example Operation of a Neural Stack



Convergence and Data Efficiency



- Where memory-enhanced models converge to perfect accuracy (most tasks), perfect accuracy is maintained on the test set (longer sequences).
- Memory-enhanced models converge order(s) of magnitude before LSTMs.
- Memory-enhanced models require relatively little data to generalise properly, vs. LSTMs, which require of the order of 1e6 minibatches of size 1000, i.e. millions of training examples, to converge on near-perfect accuracy.

Memory enhanced models learn *algorithmic* behaviour which generalises beyond the training sample.

Transduction Task Accuracies

- Separate models trained for each task.
- Evaluated on held-out sequences from training (sequence length range: 8–64) and test (sequence length range: 65–128) data sources.
- Picking the wrong memory reduces memory-enhanced LSTM to LSTM.
- Memory-enhanced models generalise better than deep LSTMs.

Experiment	Model	Training	Testing
Sequence Copying $\langle s\rangle a_1 \ldots a_k \rightarrow a_1 \ldots a_k \langle/s\rangle$	4-layer LSTM	0.98	0.50
	Stack-LSTM	0.94	0.22
	Queue-LSTM	1.00	1.00
	DeQue-LSTM	1.00	1.00
Sequence Reversal $\langle s\rangle a_1 a_2 \ldots a_k \rightarrow a_k \ldots a_2 a_1 \langle/s\rangle$	8-layer LSTM	0.98	0.13
	Stack-LSTM	1.00	1.00
	Queue-LSTM	0.61	0.01
	DeQue-LSTM	1.00	1.00
Bigram Flipping $\langle s\rangle a_1 a_2 a_3 a_4 \ldots a_{k-1} a_k \rightarrow a_2 a_1 a_4 a_3 \ldots a_k a_{k-1} \langle/s\rangle$	2-layer LSTM	0.93	0.52
	Stack-LSTM	0.90	0.48
	Queue-LSTM	0.94	0.98
	DeQue-LSTM	0.94	0.98
Subject-Verb-Object to Subject-Object-Verb (+ rel. clauses) si1 vi28 oi5 oi7 si15 rpi si19 vi16 oi10 oi24 → so1 oo5 oo7 so15 rpo so19 vo16 oo10 oo24 vo28	8-layer LSTM	0.99	0.99
	Stack-LSTM	1.00	1.00
	Queue-LSTM	1.00	1.00
	DeQue-LSTM	1.00	1.00
Pseudo English-to-German Gender Conjugation we11 the en19 and the em17 → wg11 das gn19 und der gm17	8-layer LSTM	0.99	0.99
	Stack-LSTM	0.97	0.97
	Queue-LSTM	1.00	1.00
	DeQue-LSTM	1.00	1.00