FEED-FORWARD NETWORKS WITH ATTENTION CAN SOLVE SOME LONG-TERM MEMORY PROBLEMS

Colin Raffel (craffel@gmail.com) and Daniel P. W. Ellis (dpwe@ee.columbia.edu)

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Columbia University
in the City of New York

We propose a simplified model of attention which is applicable to feed-forward neural networks and demonstrate that the resulting model can solve the synthetic "addition" and "multiplication" long-term memory problems for sequence lengths which are both longer and more widely varying than the best published results for these tasks.

Toy long-term memory problems

Input										Target	
0.5	-0.7	0.3	0.1	-0.2	...	-0.5	0.9	...	0.8	0.2	+	×
-1	0	1	0	0	...	0	1	...	0	-1	0.8	0.27

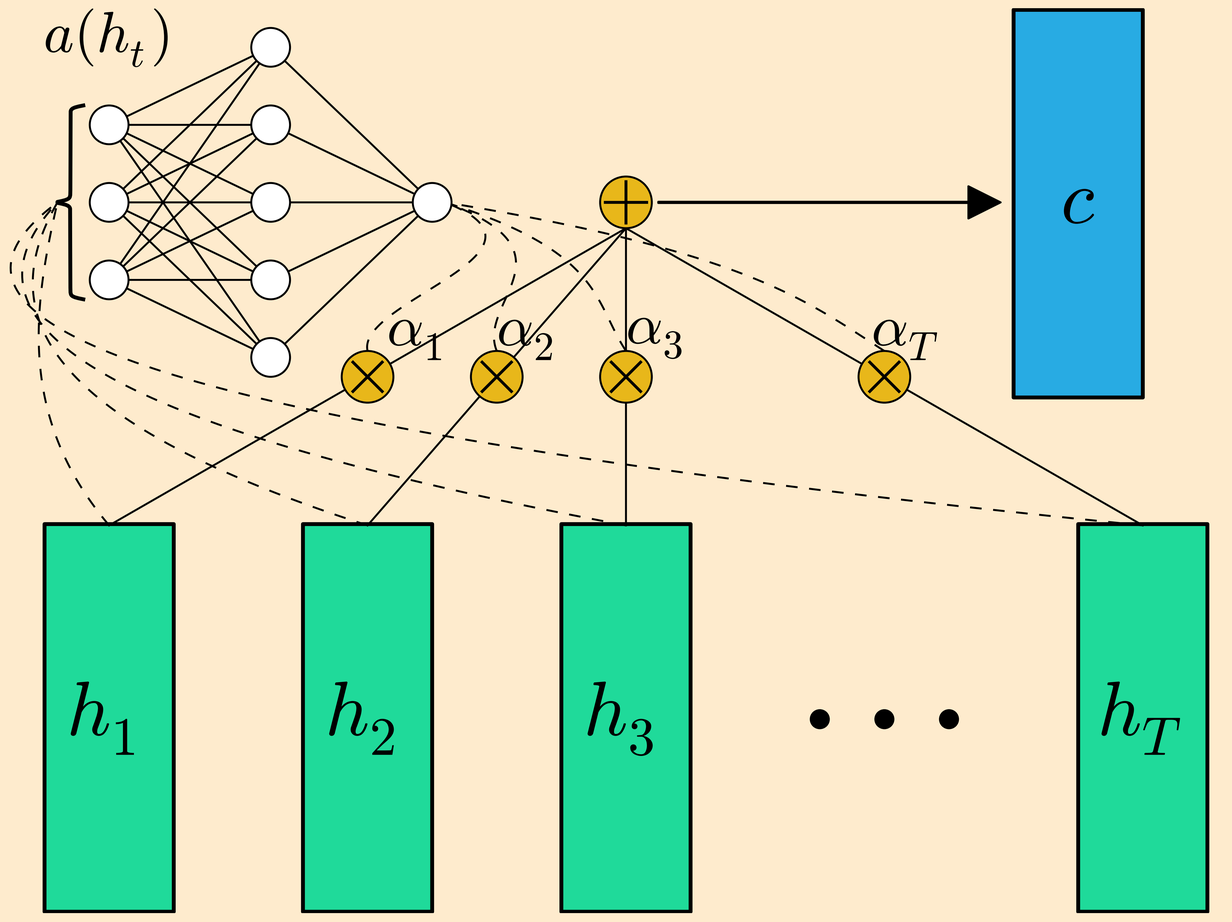
Epochs until perfect accuracy, or accuracy after 100 epochs

Task	Addition						
Length	50	100	500	1000	5000	10000	50-10000
Attention	1	1	1	1	2	3	99.9%
Unweighted	1	1	1	2	8	17	77.4%
Task	**Multiplication**						
Length	50	100	500	1000	5000	10000	50-10000
Attention	1	2	4	2	15	6	99.4%
Unweighted	2	2	8	33	89.8%	80.8%	55.5%

See also

"Pruning Subsequence Search with Attention-Based Embedding", in *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016.

Feed-forward attention



$$e_t = a(h_t) \qquad \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \qquad c = \sum_{t=1}^{T} \alpha_t h_t$$