

A. Comparison with a Neural Turing Machine

Machine

We have run additional experiments to compare Associative LSTM with a Neural Turing Machine (NTM). The comparison was done on the XML, variable assignment and arithmetic tasks. The same network architecture was used for all three tasks. The network sizes can be seen in Table 1.

The learning curves are shown in Figures 1-3. Training was done with minibatches of size 1 to be able to compare with the original Neural Turing Machine, but other minibatch sizes lead to similar learning curves. Both Associative LSTM and the Neural Turing Machine achieved good performance on the given tasks. Associative LSTM has more stable learning progress. On the other hand, the Neural Turing Machine has shown previously better generalization to longer sequences on algorithmic tasks.

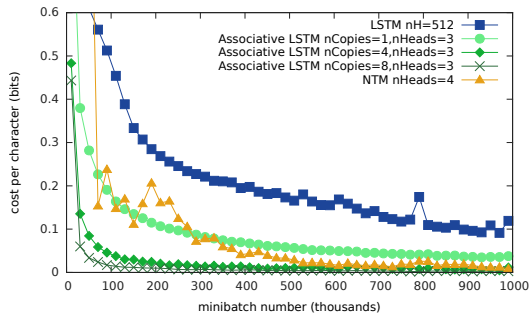


Figure 1. Training cost on the XML task.

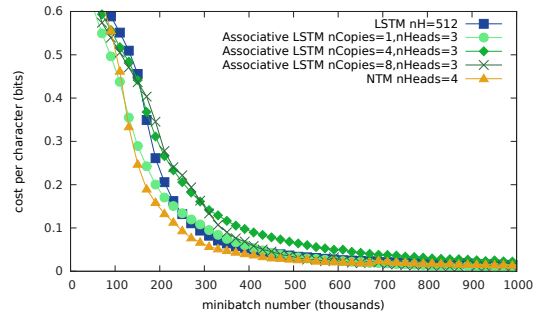


Figure 3. Training cost on the arithmetic task.

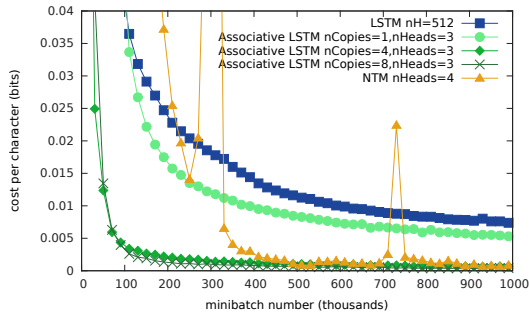


Figure 2. Training cost on the variable assignment task.

Table 1. Networks compared to a Neural Turing Machine.

Network	Memory Size	Relative speed	#parameters
LSTM nH=512	$N_h = 512$	1	1, 256, 691
Associative LSTM nHeads=3	$N_h = 128$ (=64 complex numbers)	0.66 ($N_{copies} = 1$)	775, 731
		0.56 ($N_{copies} = 4$)	
		0.46 ($N_{copies} = 8$)	
NTM nHeads=4	$N_h = 384$, memorySize = 128×20	0.66	1, 097, 340