Dynamic Neural Turing Machines with Soft and Hard Addressing Schemes

Caglar Gulcehre*, Sarath Chandar*, Kyunghyun Cho†, Yoshua Bengio*

*University of Montreal, †New York University

Motivation

- AA
- BB

Our Contributions

- We propose a generalization of Neural Turing Machine called a dynamic neural Turing machine (D-NTM) which employs learnable, location-based addressing.
- We demonstrate the application of neural Turing machines for a more complicated real task: episodic question-answering.
- We propose to use the hard attention mechanism and empirically show that it outperforms the soft attention based addressing.
- We propose a curriculum strategy for our model with the feedforward controller and discrete attention that improves our results significantly.

Neural Turing Machines

Controller

$$\mathbf{h}^t = \mathrm{GRU}(\mathbf{x}^t, \mathbf{h}^{t-1}, \mathbf{m}^t) \tag{1}$$

or for a feedforward-controller

$$\mathbf{h}^t = \sigma(\mathbf{x}^t, \mathbf{m}^t). \tag{2}$$

Memory Addressing

Reading

$$\mathbf{m}^t = \mathbf{w}^t \mathbf{M}^{t-1}, \tag{3}$$

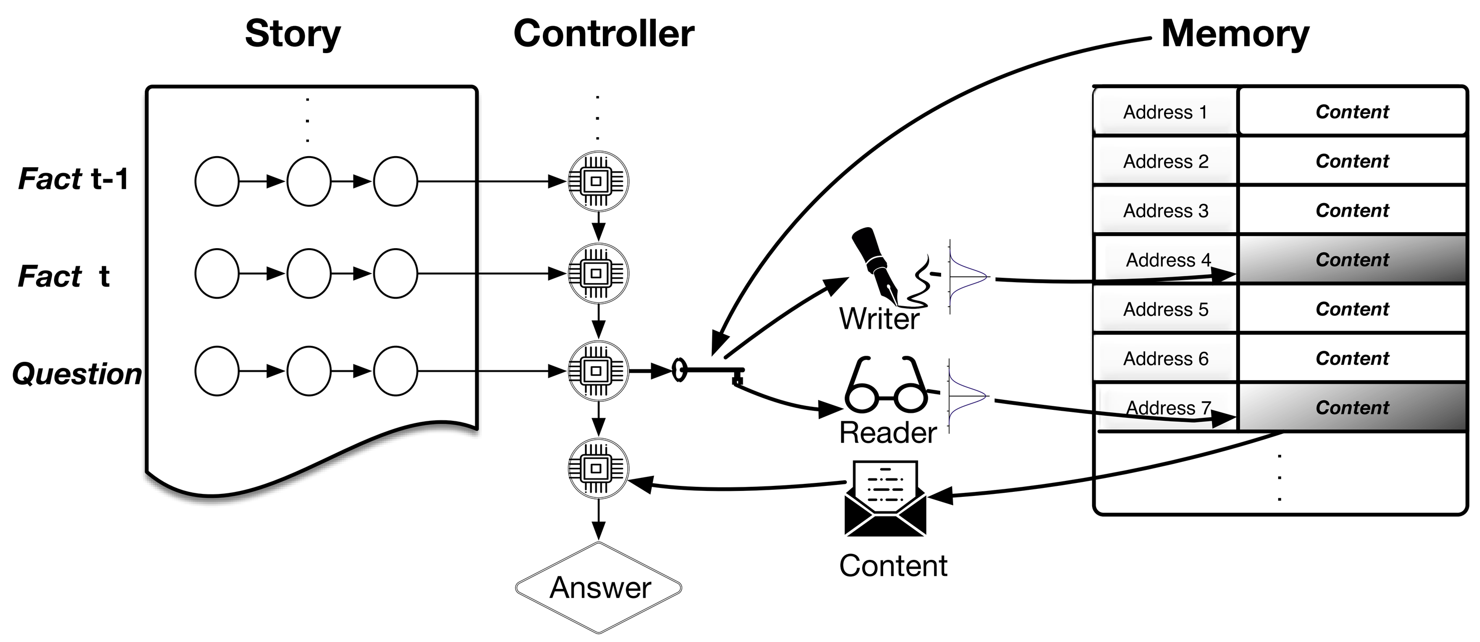
Erasing and Writing

$$\mathbf{m}_j^t = (1 - e_j^t u_j^t)\mathbf{m}_j^{t-1} + u_j^t \mathbf{c}^t, \tag{4}$$

Learning

$$C(\theta) = \frac{1}{N}\sum_{n=1}^{N} -\log p(\mathbf{y}^n | \mathbf{x}_1^n, \ldots, \mathbf{x}_T^n), \tag{5}$$

Dynamic Neural Turing Machines



A graphical illustration of the proposed dynamic neural Turing machine with the recurrent-controller. The controller receives the fact as a continuous vector encoded by a recurrent neural network, computes the read and write weights for addressing the memory. If the D-NTM automatically detects that a query has been received, it returns an answer and terminates.

Memory

$$\mathbf{M} = [\mathbf{A}; \mathbf{C}].$$

$$\mathbf{m}_i = [\mathbf{a}_i; \mathbf{c}_i].$$

No Operation (NOP) As found in **?**, an additional NOP action might be beneficial for the controller *not* to access the memory once in a while.

Address vectors First, the controller computes a key vector:

$$\mathbf{k}^t = \mathbf{W}_k^\top \mathbf{h}^t + \mathbf{b}_k^t,$$

$$\beta_t = \mathrm{softplus}(\mathbf{u}_\beta^\top \mathbf{h}^t + b_\beta).$$

$\mathbf{u}_\beta$ and b_β are similarly the head parameters.
The address vector is then computed by

$$z_i^t = \beta^t K\left(\mathbf{k}^t, \mathbf{m}_i^t\right) \tag{6}$$

$$w_i^t = \frac{\exp(z_i^t)}{\sum_j \exp(z_j^t)}, \tag{7}$$

where the similarity function K is defined as

$$K(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{(||\mathbf{x}||||\mathbf{y}|| + \epsilon)}.$$

Experiments

GRU Controller

Task	LSTM	MemN2N	DMN+	1-step LBA NTM	1-step CBA NTM	1-step Soft D-NTM	1-step Discrete D-NTM	3-steps LBA NTM	3-steps CBA NTM	3-steps Soft D-NTM	3-steps Discrete D-NTM
1	0.00	0.00	0.00	16.30	16.88	5.41	6.66	0.00	0.00	0.00	0.00
2	81.90	0.30	0.30	57.08	55.70	58.54	56.04	61.67	59.38	46.66	62.29
3	83.10	2.10	1.10	74.16	55.00	74.58	72.08	83.54	65.21	47.08	41.45
4	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	1.20	0.80	0.50	1.46	20.41	1.66	1.04	0.83	1.46	1.25	1.45
6	51.80	0.10	0.00	23.33	21.04	40.20	44.79	48.13	54.80	20.62	11.04
7	24.90	2.00	2.40	21.67	21.67	19.16	19.58	7.92	37.70	7.29	5.62
8	34.10	0.90	0.00	25.76	21.05	12.58	18.46	25.38	8.82	11.02	0.74
9	20.20	0.30	0.00	24.79	24.17	36.66	34.37	37.80	0.00	39.37	32.50
10	30.10	0.00	0.00	41.46	33.13	52.29	50.83	56.25	23.75	20.00	20.83
11	10.30	0.10	0.00	18.96	31.88	31.45	4.16	3.96	0.28	30.62	16.87
12	23.40	0.00	0.00	25.83	30.00	7.70	6.66	28.75	23.75	5.41	4.58
13	6.10	0.00	0.00	6.67	5.63	5.62	2.29	5.83	83.13	7.91	5.00
14	81.00	0.10	0.20	58.54	59.17	60.00	63.75	61.88	57.71	58.12	60.20
15	78.70	0.00	0.00	36.46	42.30	36.87	39.27	35.62	21.88	36.04	40.26
16	51.90	51.80	45.30	71.15	71.15	49.16	51.35	46.15	50.00	46.04	45.41
17	50.10	18.60	4.20	43.75	43.75	17.91	16.04	43.75	56.25	21.25	9.16
18	6.80	5.30	2.10	3.96	47.50	3.95	3.54	47.50	47.50	6.87	1.66
19	90.30	2.30	0.00	75.89	71.51	73.74	64.63	61.56	63.65	75.88	76.66
20	2.10	0.00	0.00	1.25	0.00	2.70	3.12	0.40	0.00	3.33	0.00
Avg.Err.	36.41	4.24	**2.81**	31.42	33.60	29.51	**27.93**	32.85	32.76	24.24	**21.79**

Feedforward Controller

Task	LSTM	MemN2N	DMN+	Soft D-NTM	Discrete D-NTM	Discrete* D-NTM
1	0.00	0.00	0.00	4.38	81.67	14.79
2	81.90	0.30	0.30	27.5	76.67	76.67
3	83.10	2.10	1.10	71.25	79.38	70.83
4	0.20	0.00	0.00	0.00	78.65	44.06
5	1.20	0.80	0.50	1.67	83.13	17.71
6	51.80	0.10	0.00	1.46	48.76	48.13
7	24.90	2.00	2.40	6.04	54.79	23.54
8	34.10	0.90	0.00	1.70	69.75	35.62
9	20.20	0.30	0.00	0.63	39.17	14.38
10	30.10	0.00	0.00	19.80	56.25	56.25
11	10.30	0.10	0.00	0.00	78.96	39.58
12	23.40	0.00	0.00	6.25	82.5	32.08
13	6.10	0.00	0.00	7.5	75.0	18.54
14	81.00	0.10	0.20	17.5	78.75	24.79
15	78.70	0.00	0.00	0.0	71.42	39.73
16	51.90	51.80	45.30	49.65	71.46	71.15
17	50.10	18.60	4.20	1.25	43.75	43.75
18	6.80	5.30	2.10	0.24	48.13	2.92
19	90.30	2.30	0.00	39.47	71.46	71.56
20	2.10	0.00	0.00	0.0	76.56	9.79
Avg.Err.	36.41	4.24	**2.81**	**12.81**	68.30	37.79