doi:10.1038/nature20101

Name	Description	Domain
t	time-step	$\mathbb{N}$
N	number of memory locations	$\mathbb{N}$
W	memory word size	$\mathbb{N}$
R	number of read heads	$\mathbb{N}$
$\mathbf{x}_t$	input vector	$\mathbb{R}^X$
$\mathbf{y}_t$	output vector	$\mathbb{R}^Y$
$\mathbf{z}_t$	target vector	$\mathbb{R}^Y$
$\mathbf{M}_t$	memory matrix	$\mathbb{R}^{N \times W}$
$\mathbf{k}_t^{r,i}$	read key i $(1 \leq i \leq R)$	$\mathbb{R}^W$
$\mathbf{r}_t^i$	read vector i	$\mathbb{R}^W$
$\beta_t^{r,i}$	read strength i	$[1, \infty)$
$\mathbf{k}_t^w$	write key	$\mathbb{R}^W$
β_t^w	write strength	$[1, \infty)$
$\mathbf{e}_t$	erase vector	$[0,1]^W$
$\mathbf{v}_t$	write vector	$\mathbb{R}^W$
f_t^i	free gate i	$[0,1]$
g_t^a	allocation gate	$[0,1]$
g_t^w	write gate	$[0,1]$
$\boldsymbol{\psi}_t$	memory retention vector	$\mathbb{R}^N$
$\mathbf{u}_t$	memory usage vector	$\mathbb{R}^N$
$\boldsymbol{\phi}_t$	indices of slots sorted by usage	$\mathbb{N}^N$
$\mathbf{a}_t$	allocation weighting	$\Delta_N = \{\boldsymbol{\alpha} \in \mathbb{R}^N : \boldsymbol{\alpha}_i \in [0,1], \sum_{i=1}^N \boldsymbol{\alpha}_i \leq 1\}$
$\mathbf{c}_t^w$	write content weighting	$\mathcal{S}_N = \{\boldsymbol{\alpha} \in \mathbb{R}^N : \boldsymbol{\alpha}_i \in [0,1], \sum_{i=1}^N \boldsymbol{\alpha}_i = 1\}$
$\mathbf{w}_t^w$	write weighting	Δ_N
$\mathbf{p}_t$	precedence weighting	Δ_N
$\mathbf{E}$	matrix of ones ($\mathbf{E}[i,j] = 1 \ \forall i,j$)	$\mathbb{R}^{N \times W}$
$\mathbf{L}_t$	temporal link matrix	$\mathbb{R}^{N \times N}$
$\mathbf{f}_t^i$	forward weighting i	Δ_N
$\mathbf{b}_t^i$	backward weighting i	Δ_N
$\mathbf{c}_t^{r,i}$	read content weighting i	$\mathcal{S}_N$
$\mathbf{w}_t^{r,i}$	read weighting i	Δ_N
$\boldsymbol{\pi}_t^i$	read mode i	$\mathcal{S}_3$
W_r	read key weights	$\mathbb{R}^{(RW) \times Y}$
$\boldsymbol{\theta}$	controller weights	$\mathbb{R}^\Theta$
$\boldsymbol{\xi}_t$	interface vector	$\mathbb{R}^{(W \times R)+3W+5R+3}$
$\boldsymbol{\chi}_t$	controller input vector	$\mathbb{R}^{(W \times R)+X}$
$\boldsymbol{\upsilon}_t$	controller output vector	$\mathbb{R}^Y$
$\mathcal{N}(.;\boldsymbol{\theta})$	controller network	$\left[\mathbb{R}^{(W \times R)+X}\right]^* \times \mathbb{R}^\Theta \mapsto \mathbb{R}^{(W \times R)+3W+5R+3} \times \mathbb{R}^Y$

Table 1: **DNC glossary**

DNC Equations

Definitions:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$
$$\text{oneplus}(x) = 1 + \log(1+e^{x})$$
$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{|\mathbf{x}|} e^{x_j}}$$
$$\mathcal{C}(\mathbf{M},\mathbf{k},\beta)[i] = \frac{\exp\left\{\mathcal{D}(\mathbf{k},\mathbf{M}[i,.])\beta\right\}}{\sum_j \exp\left\{\mathcal{D}(\mathbf{k},\mathbf{M}[j,.])\beta\right\}}$$
$$\mathcal{D}(\mathbf{u},\mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}$$
$$(\mathbf{A}\circ\mathbf{B})[i,j] = \mathbf{A}[i,j]\mathbf{B}[i,j];\ (\mathbf{x}\circ\mathbf{y})[i] = \mathbf{x}[i]\mathbf{y}[i]$$

Initial Conditions:

$$\mathbf{u}_0 = \mathbf{0};\ \mathbf{p}_0 = \mathbf{0};\ \mathbf{L}_0 = \mathbf{0};\ \mathbf{L}_t[i,i] = 0\ \forall i$$

Controller Update:

$$\boldsymbol{\chi}_t = [\mathbf{x}_t;\mathbf{r}_{t-1}^1;\ldots;\mathbf{r}_{t-1}^R]$$
$$(\boldsymbol{\xi}_t,\boldsymbol{\upsilon}_t) = \mathcal{N}([\boldsymbol{\chi}_1;\ldots;\boldsymbol{\chi}_t];\boldsymbol{\theta})$$

Interface Variables:

$$\boldsymbol{\xi}_t = [\mathbf{k}_t^{r,1};\ldots;\mathbf{k}_t^{r,R};\hat{\beta}_t^{r,1};\ldots,\hat{\beta}_t^{r,R};\mathbf{k}_t^w;\hat{\beta}_t^w;\hat{\mathbf{e}}_t;\mathbf{v}_t;\hat{f}_t^1;\ldots;\hat{f}_t^R;\hat{g}_t^a;\hat{g}_t^w;\hat{\boldsymbol{\pi}}_t^1;\ldots;\hat{\boldsymbol{\pi}}_t^R]$$
$$\beta_t^{r,i} = \text{oneplus}(\hat{\beta}_t^{r,i});\ \beta_t^w = \text{oneplus}(\hat{\beta}_t^w);\ \mathbf{e}_t = \sigma(\hat{\mathbf{e}}_t);\ f_t^i = \sigma(\hat{f}_t^i)$$
$$g_t^a = \sigma(\hat{g}_t^a);\ g_t^w = \sigma(\hat{g}_t^w);\ \boldsymbol{\pi}_t^k = \text{softmax}(\hat{\boldsymbol{\pi}}_t^k)$$

Memory Updates:

$$\boldsymbol{\psi}_t = \prod_{i=1}^{R}(\mathbf{1} - f_t^i\mathbf{w}_{t-1}^{r,i})$$
$$\mathbf{u}_t = \left(\mathbf{u}_{t-1} + \mathbf{w}_{t-1}^w - (\mathbf{u}_{t-1}\circ\mathbf{w}_{t-1}^w)\right)\circ\boldsymbol{\psi}_t$$
$$\boldsymbol{\phi}_t = \text{SortIndicesAscending}(\mathbf{u}_t)$$
$$\mathbf{a}_t[\boldsymbol{\phi}_t[j]] = (1 - \mathbf{u}_t[\boldsymbol{\phi}_t[j]])\prod_{i=1}^{j-1}\mathbf{u}_t[\boldsymbol{\phi}_t[i]]$$
$$\mathbf{c}_t^w = \mathcal{C}(\mathbf{M}_{t-1},\mathbf{k}_t^w,\beta_t^w)$$
$$\mathbf{w}_t^w = g_t^w\left[g_t^a\mathbf{a}_t + \left(1-g_t^a\right)\mathbf{c}_t^w\right]$$
$$\mathbf{M}_t = \mathbf{M}_{t-1}\circ(\mathbf{E} - \mathbf{w}_t^w\mathbf{e}_t^\top) + \mathbf{w}_t^w\mathbf{v}_t{}^\top$$
$$\mathbf{p}_t = \left(1 - \sum_i \mathbf{w}_t^w[i]\right)\mathbf{p}_{t-1} + \mathbf{w}_t^w$$
$$\mathbf{L}_t[i,j] = \left(1 - \mathbf{w}_t^w[i] - \mathbf{w}_t^w[j]\right)\mathbf{L}_{t-1}[i,j] + \mathbf{w}_t^w[i]\mathbf{p}_{t-1}[j]$$
$$\mathbf{f}_t^i = \mathbf{L}_t\mathbf{w}_{t-1}^{r,i}$$
$$\mathbf{b}_t^i = \mathbf{L}_t^\top\mathbf{w}_{t-1}^{r,i}$$
$$\mathbf{c}_t^{r,i} = \mathcal{C}(\mathbf{M}_t,\mathbf{k}_t^{r,i},\beta_t^{r,i})$$
$$\mathbf{w}_t^{r,i} = \boldsymbol{\pi}_t^i[1]\mathbf{b}_t^i + \boldsymbol{\pi}_t^i[2]\mathbf{c}_t^{r,i} + \boldsymbol{\pi}_t^i[3]\mathbf{f}_t^i$$
$$\mathbf{r}_t^i = \mathbf{M}_t^\top\mathbf{w}_t^{r,i}$$

Output:

$$\mathbf{y}_t = W_r[\mathbf{r}_t^1;\ldots;\mathbf{r}_t^R] + \boldsymbol{\upsilon}_t$$