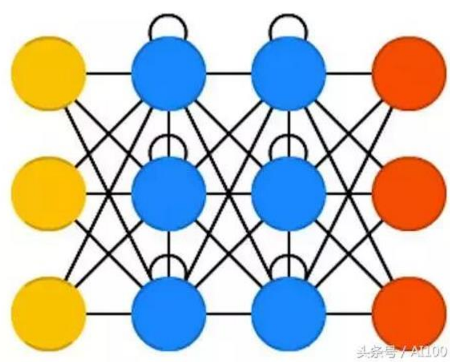
# Neural Turing machines

Tim 16th July

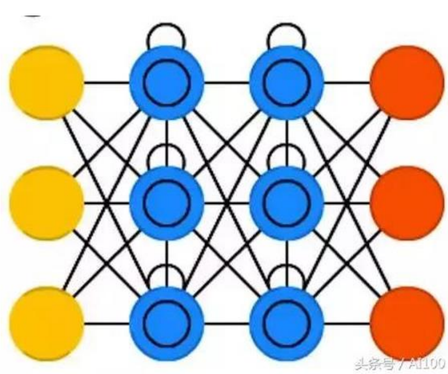
## CONTENTS

# RNN

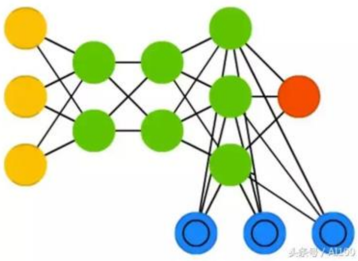



Problem:
Input order will affect the training result of neural network.
The gradient disappears (or the gradient explodes, depending on the activation function used), and the information disappears quickly over time

# LSTM



# NTM



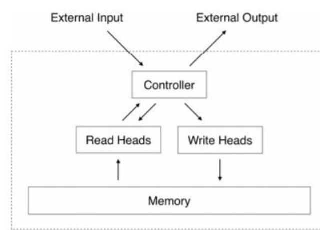



**Figure 1: Neural Turing Machine Architecture.** During each update cycle, the controller network receives inputs from an external environment and emits outputs in response. It also reads to and writes from a memory matrix via a set of parallel read and write heads. The dashed line indicates the division between the NTM circuit and the outside world.

2013) (Bahdanau et al., 2014) and program search (Hochreiter et al., 2001b) (Das et al., 1992), constructed with recurrent neural networks.

# Read

$$\sum_i w_t(i) = 1, \qquad 0 \le w_t(i) \le 1, \ \forall i. \tag{1}$$

$$\mathbf{r}_t \longleftarrow \sum_i w_t(i)\mathbf{M}_t(i), \tag{2}$$

# Write

$$\tilde{\mathbf{M}}_t(i) \longleftarrow \mathbf{M}_{t-1}(i)\left[\mathbf{1} - w_t(i)\mathbf{e}_t\right], \qquad (3)$$

$$\mathbf{M}_t(i) \longleftarrow \tilde{\mathbf{M}}_t(i) + w_t(i)\,\mathbf{a}_t. \qquad (4)$$

# Addressing Mechanisms

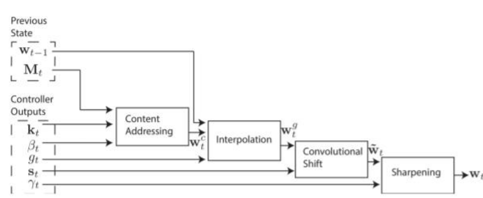



**Figure 2: Flow Diagram of the Addressing Mechanism.** The *key vector*, $\mathbf{k}_t$, and *key strength*, $\beta_t$, are used to perform content-based addressing of the memory matrix, $\mathbf{M}_t$. The resulting content-based weighting is interpolated with the weighting from the previous time step based on the value of the *interpolation gate*, $g_t$. The *shift weighting*, $\mathbf{s}_t$, determines whether and by how much the weighting is rotated. Finally, depending on $\gamma_t$, the weighting is sharpened and used for memory access.

# Focusing by Content

$$w_t^c(i) \longleftarrow \frac{\exp\Big(\beta_t K\big[\mathbf{k}_t, \mathbf{M}_t(i)\big]\Big)}{\sum_j \exp\Big(\beta_t K\big[\mathbf{k}_t, \mathbf{M}_t(j)\big]\Big)}. \tag{5}$$

$$K\big[\mathbf{u}, \mathbf{v}\big] = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \cdot ||\mathbf{v}||}. \tag{6}$$

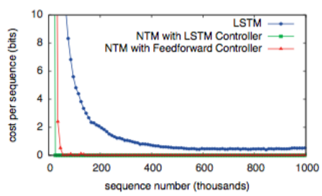
# Focusing by Location

$$\mathbf{w}_t^g \longleftarrow g_t \mathbf{w}_t^c + (1 - g_t)\mathbf{w}_{t-1}. \tag{7}$$

$$\tilde{w}_t(i) \longleftarrow \sum_{j=0}^{N-1} w_t^g(j)\, s_t(i-j) \tag{8}$$

$$w_t(i) \longleftarrow \frac{\tilde{w}_t(i)^{\gamma_t}}{\sum_j \tilde{w}_t(j)^{\gamma_t}} \tag{9}$$

# Experiments

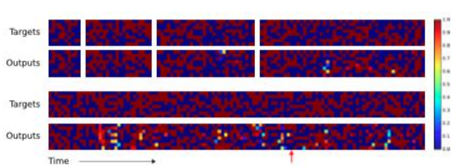


Copy Learning Curves.

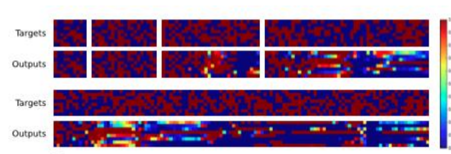
```
initialise: move head to start location
while input delimiter not seen do
    receive input vector
    write input to head location
    increment head location by 1
end while
return head to start location
while true do
    read output vector from head location
    emit output
    increment head location by 1
end while
```

pseudocode

# Experiments



NTM



LSTM

## Conclusion

NTM is fast and efficient, which has the abilities of creation and iteration.

Formula 7 represents the moving depending on the last moment, which means that the long series can be computed.

Formula 9 represents the protection of the weight, as the time goes by.