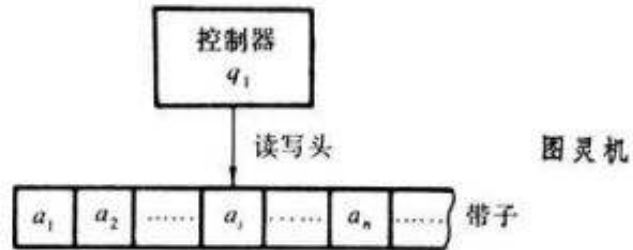

Neural Turing Machines

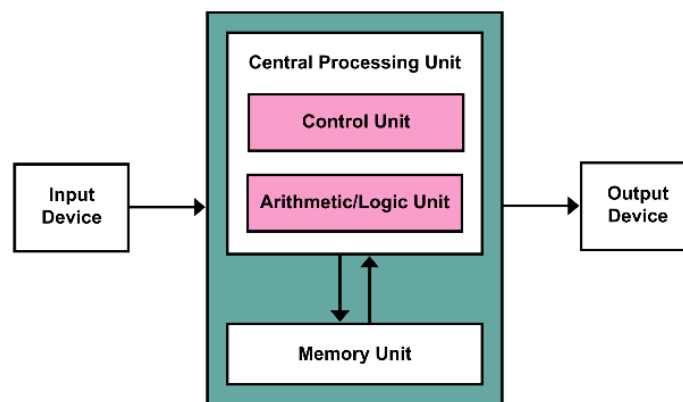
<https://www.dosrc.com/>

1 相关知识

图灵机:图灵机是一个理论的计算模型，由计算机技术先驱Turing在1936年提出



冯·诺伊曼体系结构:

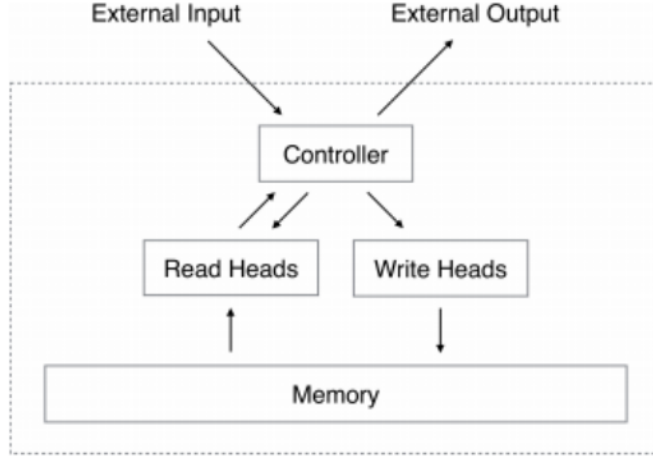


2 神经网络图灵机

计算机程序在执行计算任务的过程中(Von Neumann, 1945)使用了三个基本机制：初等运算（如算术操作），逻辑控制流（分支循环）和可读写的存储器。虽然现代机器学习理论在建模复杂数据方面取得了广泛的成功，却普遍忽略了对控制流和存储器的使用。本文通过

引入外部存储器（external memory）来增强神经网络的能力，神经网络和外部存储器之间通过注意力机制进行交互。新系统可以与图灵机或者冯·诺依曼体系相类比，但每个组成部分都是可微的，可以使用梯度下降进行高效训练。初步的结果显示神经网络图灵机能够从输入和输出样本中推理出简单的算法，如复制、排序和关联性回忆。

2.1 网络结构



2.2 存储器结构

存储器是一个 N 行 M 列的矩阵

2.3 读存储器操作

$$\sum_i w_t(i) = 1, \quad 0 \leq w_t(i) \leq 1, \quad \forall i.$$

$$r_t \leftarrow \sum_i w_t(i) M_t(i)$$

其中 \mathbf{w}_t 是一个 N 维向量，代表读权重

2.4 写存储器操作

$$\tilde{\mathbf{M}}_t(i) \leftarrow \mathbf{M}_{t-1}(i) [\mathbf{1} - w_t(i) \mathbf{e}_t]$$

$$\mathbf{M}_t(i) \leftarrow \tilde{\mathbf{M}}_t(i) + w_t(i) \mathbf{a}_t$$

其中 \mathbf{w}_t 是一个 N 维向量，代表写权重

其中 \mathbf{e}_t 是一个 M 维向量，代表擦除权重

其中 \mathbf{a}_t 是一个 M 维向量，代表写入的值向量

2.5 寻址方式

2.5.1 内容寻址

即按内容的相关性来寻址，读写头先产生一个 \mathbf{k}_t 的键值，通过与每个内存中的值 $\mathbf{M}_t(i)$ 相比较，来产生一个 N 维的相似度向量 \mathbf{w}_t^c ，其表达式如下：

$$w_t^c(i) \leftarrow \frac{\exp(\beta_t K[\mathbf{k}_t, \mathbf{M}_t(i)])}{\sum_j \exp(\beta_t K[\mathbf{k}_t, \mathbf{M}_t(j)])}$$

在本文中相似度函数 $K[\cdot, \cdot]$ 由余弦相似度给出，如下：

$$K[\mathbf{u}, \mathbf{v}] = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

其中 β_t 表示注意力集中的精度

2.5.2 位置寻址

1. 插值

$$\mathbf{w}_t^g \leftarrow g_t \mathbf{w}_t^c + (1 - g_t) \mathbf{w}_{t-1}$$

其中 g_t 表示插值的一个门，可理解为一个控制遗忘程度的权重。 w_t^c 是由内容寻址产生的权重值， w_{t-1} 是前一时刻的寻址权重值。

2. 循环卷积

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

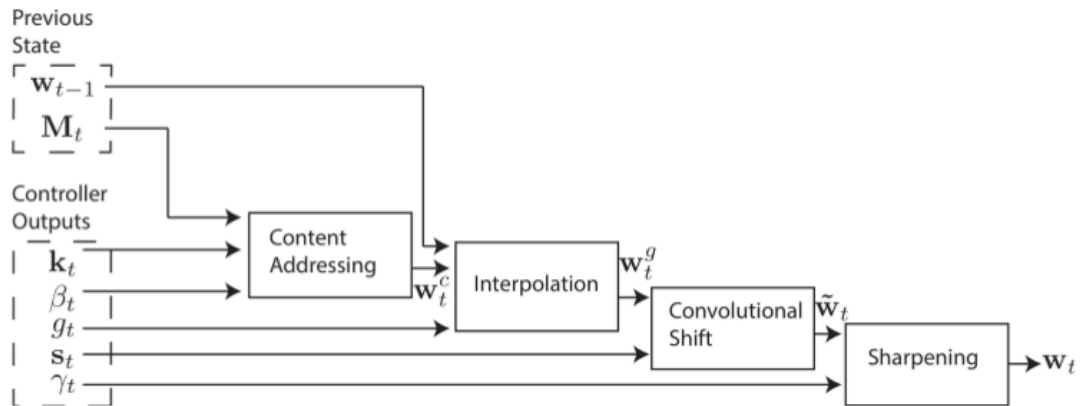
其中 s_t 代表循环卷积的向量。循环卷积是为了移动读写头的位置。

2. 锐化

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

其中标量 γ_t 代表锐化的程度，并且 $\gamma_t \geq 1$ 。锐化是防止循环卷积后权重值的分散。

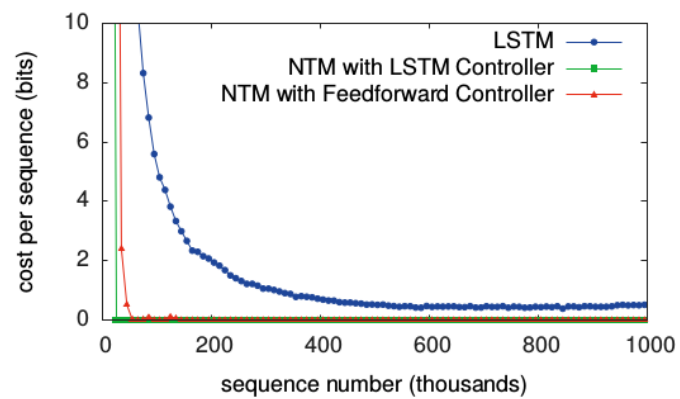
2.5.3 寻址流程图



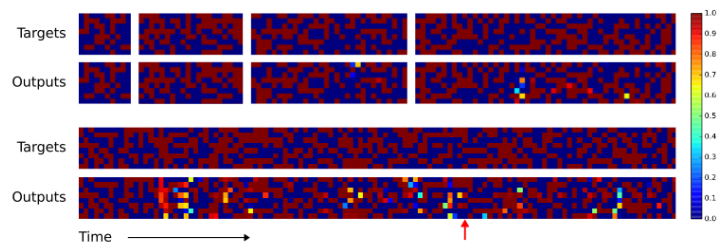
3 实验

3.1 复制任务

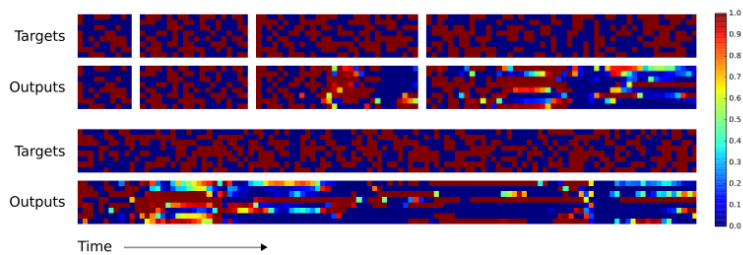
复制任务的学习训练曲线:



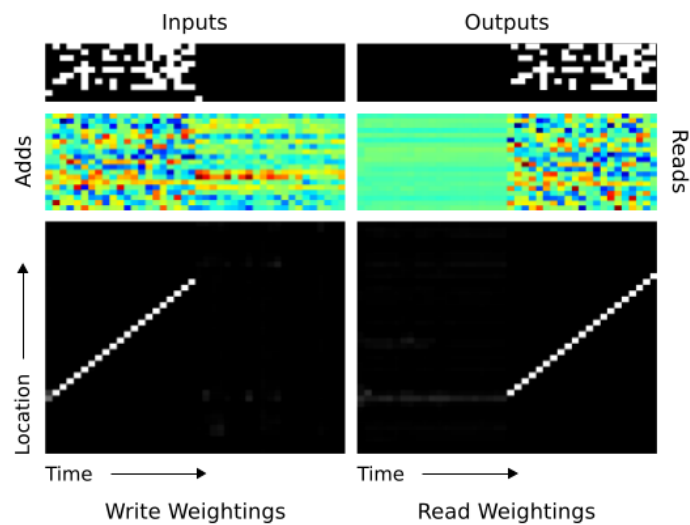
NMT在复制任务上的泛化能力:



LSTM在复制任务上的泛化能力:



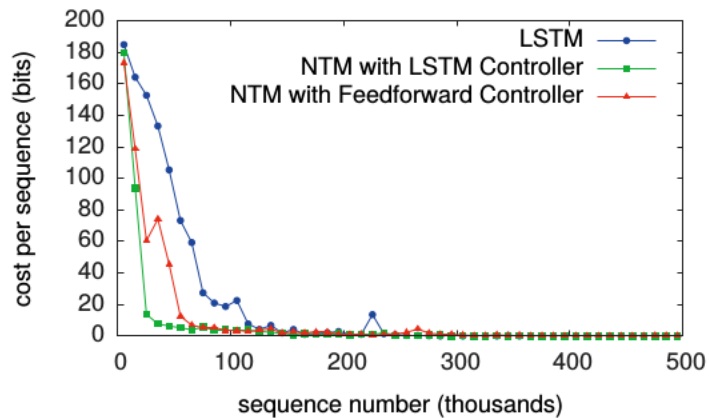
NMT在复制任务上的存储器使用情况:



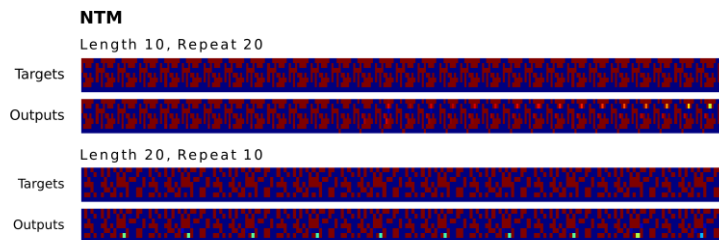
可以看出基本和低级语言编写的程序存储器使用情况一致。

3.2 重复复制任务

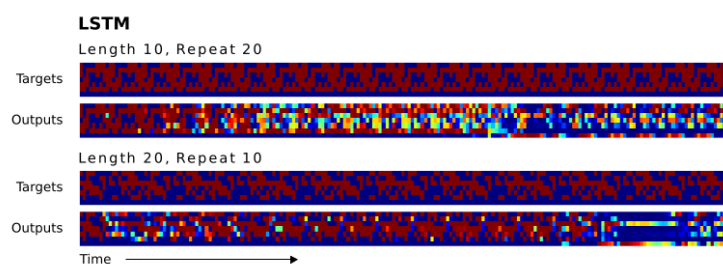
重复复制任务的学习训练曲线:



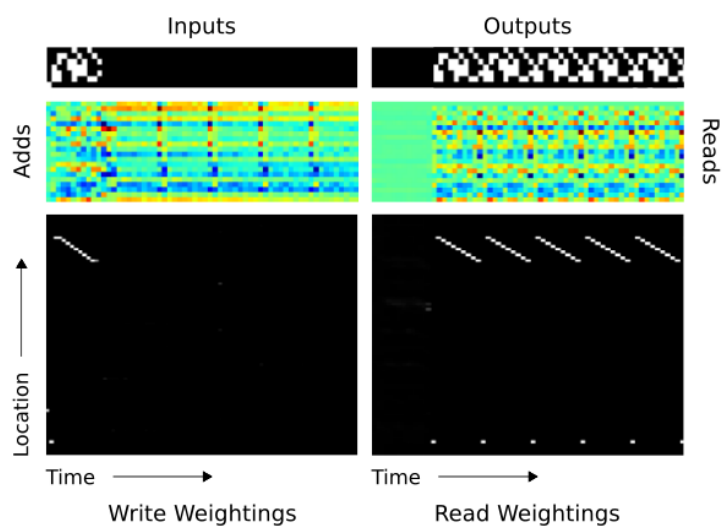
NMT在重复复制任务上的泛化能力:



LSTM在重复复制任务上的泛化能力:



NMT在重复复制任务上的存储器使用情况:



可以看出基本和低级语言编写的程序存储器使用情况一致。