

## 预训练模型学习情况周报 2

姚凯

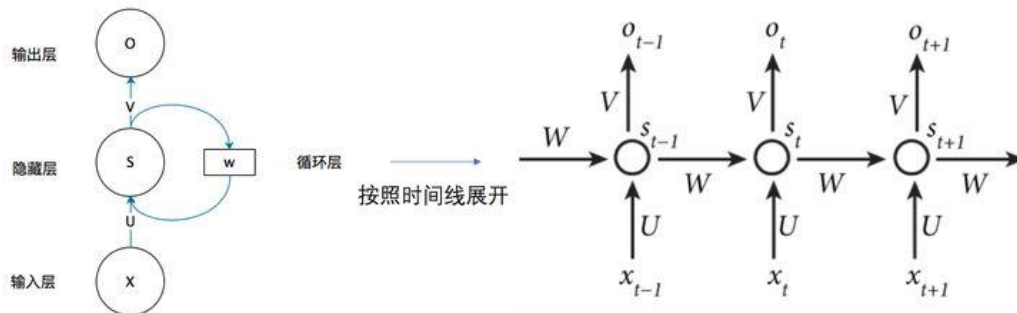
### 一、本周学习：

#### 《神经网络与深度学习》

#### 第六章：循环神经网络

处理网络输出与输入有关，或者变长时序数据的问题

简单循环网络



$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$

#### 第七章：网络优化与正则化

网络优化：找到全局最优，克服梯度消失和梯度爆炸

调整学习率和梯度估计修正、参数初始化方法（基于固定方差、方差缩放）、数

据预处理方法、逐层归一化方法和超参数优化方法（网格搜索，随机搜索）

网络正则化：避免过拟合

$\ell_1$  和  $\ell_2$  正则化、权重衰减、提前停止、丢弃法、数据增强和标签平滑

#### 第八章：注意力机制与外部记忆

提高神经网络处理信息的能力

1) 注意力：通过自上而下的信息选择机制来过滤掉大量的无关信息

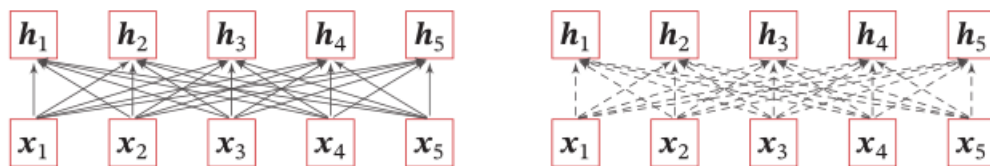
注意力分布：

$$\begin{aligned}\alpha_n &= p(z = n | \mathbf{X}, \mathbf{q}) \\ &= \text{softmax}(s(\mathbf{x}_n, \mathbf{q})) \\ &= \frac{\exp(s(\mathbf{x}_n, \mathbf{q}))}{\sum_{j=1}^N \exp(s(\mathbf{x}_j, \mathbf{q}))},\end{aligned}$$

加权平均：

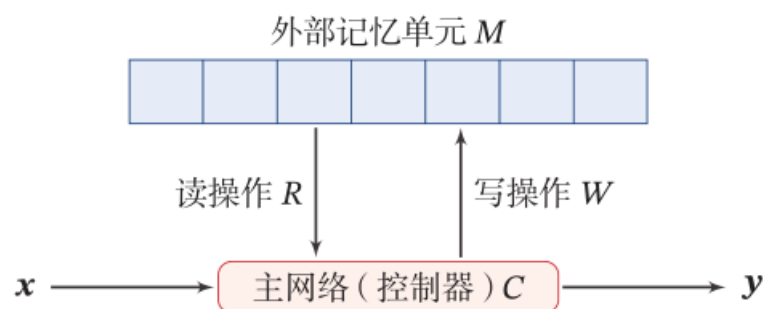
$$\begin{aligned}\text{att}(\mathbf{X}, \mathbf{q}) &= \sum_{n=1}^N \alpha_n \mathbf{x}_n, \\ &= \mathbb{E}_{z \sim p(z | \mathbf{X}, \mathbf{q})}[\mathbf{x}_z].\end{aligned}$$

自注意力模型：



自注意力模型的权重是动态生成的，因此可以处理变长的信息序列。

2) 外部记忆：辅助记忆单元以提高神经网络存储信息的容量



主要有端到端记忆网络，神经图灵机和 Hopfield 网络等

二、下周学习：

基本看完自然语言处理的深度学习基础 ppt, 以及 Pytorch 官方教程 Introduction to Pytorch 部分