

更多神经网络结构

该课程主要为大家讲授如下的内容：

- CNN
- 循环神经网络
- LSTM
- GRU
- 注意力机制
- 自注意力机制
- Transformer

1. 典型的神经网络结构

1. CNN

卷积神经网络（Convolutional Neural Network, CNN）是通过对输入信号进行卷积计算而提取特征的一种神经网络，主要用于图像处理。

卷积神经网络

Convolutional Neural Network, CNN

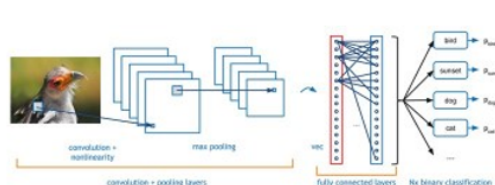
Microsoft AI Talent Program

卷积层

卷积运算

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		



(c)Microsoft 2022. 仅供个人学习使用

2. 循环神经网络

循环神经网络（Recurrent Neural Network, RNN）是一种引入时

态叠加、为当前输出添加上一时态参数的网络结构。

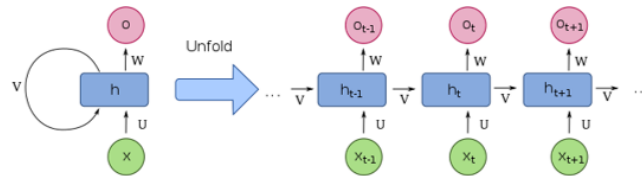
循环神经网络

Recurrent Neural Network, RNN

Microsoft AI Talent Program

神经元接受自身的状态信息作为输入

其状态信息在网络中循环传递



(c)Microsoft 2022. 仅供个人学习使用

3. LSTM

长短时记忆 (Long-short Term Memory, LSTM) 是基于RNN的设计发展而来的一种模型。它对单个神经元进行了更复杂的设计, 添加输入门、输出门和遗忘门, 使单个神经元能够叠加更多时态的影响。LSTM一度被广泛用于序列处理当中。

LSTM由德国学者 Hochreiter和Schmidhuber提出于1997年。

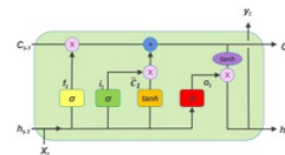
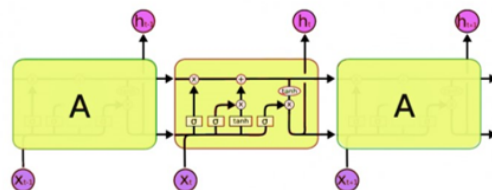
长短时记忆

(Long Short Term Memory, LSTM)

Microsoft AI Talent Program

一种神经元更加复杂的 RNN

处理时间序列中当间隔和延迟较长时



(c)Microsoft 2022. 仅供个人学习使用

4. GRU

门循环单元 (Gate Recurrent Unit, GRU) 是一种和LSTM设计相仿的模型。它对单个神经元减少了计算单元、也因此减少了参数, 计算更快, 但是在实践中也能达到和LSTM同样的效果。

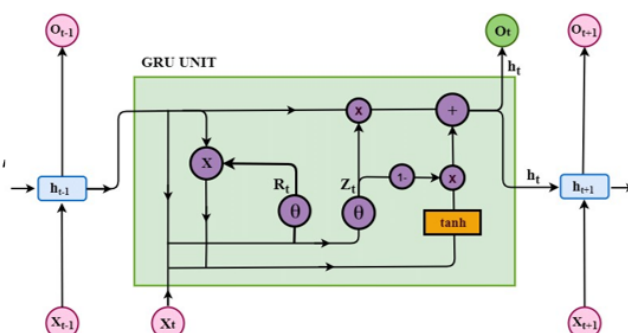
GRU在2014年由Bengio 团队的Cho等人提出。

门循环单元

(Gate Recurrent Unit, GRU)

Microsoft AI Talent Program

- LSTM的变体
- 较LSTM网络的结构更加简单,且效果也很好
 - 比LSTM少了一个“门控”
 - 参数比LSTM少



(c)Microsoft 2022. 仅供个人学习使用

5. 注意力机制 (Attention Mechanism)

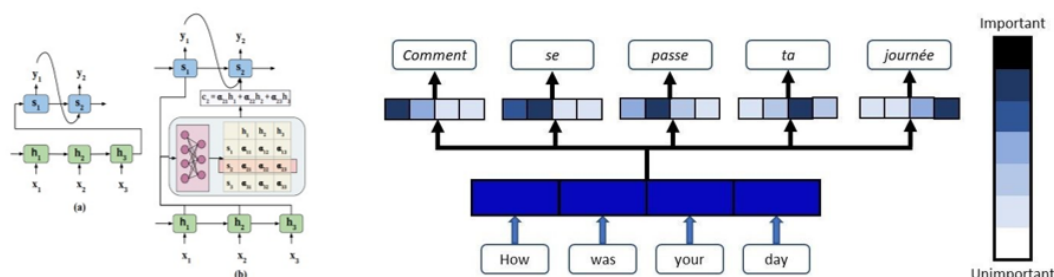
注意力机制 (Attention Mechanism) 是一种对输入信息加权的模型/层, 是一种计算能力有限情况下的资源分配方案。它在有限计算资源的条件下达到加大关键信息的权重、裁剪次要或无关信息的效果, 提高模型预测的精确度、减少算力和存储的开销。现已广泛用于机器翻译、语音识别、图像标注等领域。

注意力机制于2014年由由Bengio团队的Bahdanau 等人提出。

注意力机制 (Attention Mechanism)

Microsoft AI Talent Program

- 对输入信息的每个部分赋予不同的权重, 抽取更加关键及重要的信息
- 且不会带来计算或存储的更大开销
- 应用于机器翻译、语音识别、图像标注等领域



(c)Microsoft 2022. 仅供个人学习使用

6. 自注意力机制

自注意力机制 (Self-attention Mechanism) 是注意力机制的变体。自注意力机制和注意力机制对输入数据进行加权操作的方式不同, 带来的效果也不一样。自注意力机制也同样广泛用于机器

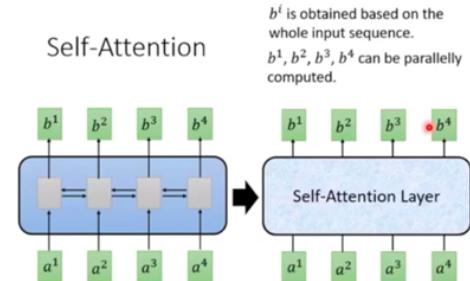
阅读、摘要总结、图像描述生成等领域。

自注意力机制于2017年由Google机器翻译团队提出。

自注意力机制 (Self-Attention Mechanism)

Microsoft AI Talent Program

- 注意力机制的变体
- 减少了对外部信息的依赖
- 捕捉数据的内部相关性
- 应用于机器阅读、摘要总结、图像描述生成等领域



(c)Microsoft 2022, 仅供个人学习使用

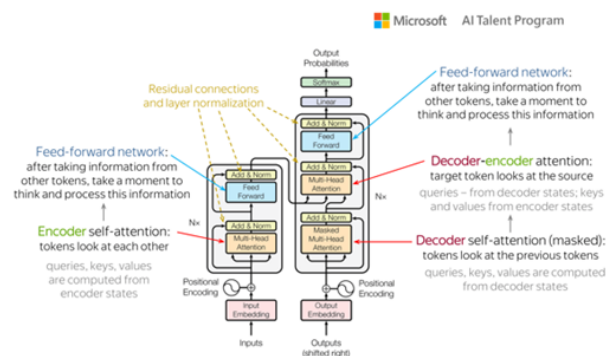
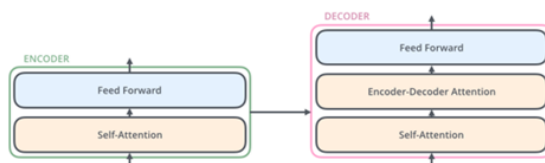
7. Transformer

Transformer是一种不同于以往CNN或RNN及其变体的模型。单个transformer层由encoder和decoder组成，每个部分又由前馈神经网络和注意力机制等构成。它本身的结构更加适用于并行计算，在精度和性能上都高于此前流行的RNN类模型。它被大量应用在翻译、文本总结等自然语言处理任务上，最近在计算机视觉上也颇有成效。

Transformer于2017年由Google Brain中的一个团队提出。

Transformer

- 由Self-Attention和前馈神经网络组成
- Encoder-Decoder架构
- 适合处理长期依赖 (long-range dependencies) 问题



Microsoft AI Talent Program