

Neural Machine Translation

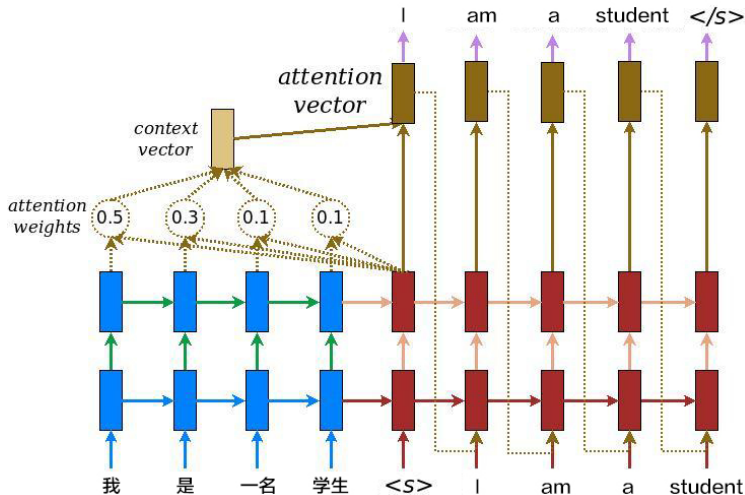
深度学习技术与应用作业 #4

王宇涛

北京大学

2020 年 5 月 10 日

网络结构



网络结构

参考 Luong 的论文，采用基于注意力的编码器 - 解码器模型，注意力机制为每个输入单词分配一个权重，然后解码器将这个权重用于预测句子中的下一个单词。下面是注意力机制的实现：

$$\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_s) = \mathbf{v}_a^\top \tanh(\mathbf{W}_1 \mathbf{h}_t + \mathbf{W}_2 \bar{\mathbf{h}}_s) \quad [\text{Attention score}]$$

$$\alpha_{ts} = \frac{\exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_s))}{\sum_{s'=1}^S \exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_{s'}))} \quad [\text{Attention weights}]$$

$$\mathbf{c}_t = \sum_s \alpha_{ts} \bar{\mathbf{h}}_s \quad [\text{Context vector}]$$

$$\mathbf{a}_t = f(\mathbf{c}_t, \mathbf{h}_t) = \tanh(\mathbf{W}_c [\mathbf{c}_t; \mathbf{h}_t]) \quad [\text{Attention vector}]$$

超参数

编码器

- ▶ Embedding - output_dim = 512
- ▶ GRU - units = 1024

解码器

- ▶ W1 W2 Dense - units = 10
- ▶ Embedding - output_dim = 512
- ▶ GRU - units = 1024

训练方法和优化方法

汉语输入

- ▶ "<start> 我 是 一 名 学 生 。 <end>"
- ▶ [1, 4, 12, 380, 248, 3, 2]

英语输出

- ▶ "<start> i am a student . <end>"
- ▶ [1, 4, 100, 8, 408, 3, 2]

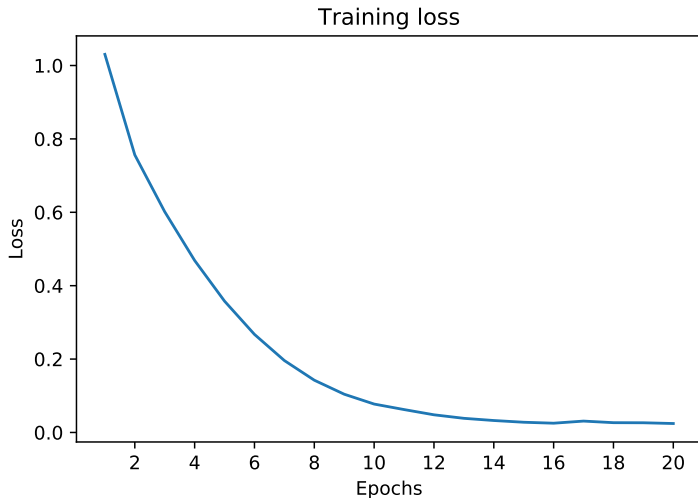
训练方法

- ▶ 训练方法为 Adam, 训练 20 轮

损失函数

- ▶ 稀疏分类交叉熵 Sparse categorical crossentropy

训练曲线



测试集评价

BLEU 指标

- ▶ BLEU-1: 0.550518 BLEU-2: 0.392867
- ▶ BLEU-3: 0.302809 BLEU-4: 0.233100

ROUGE 指标

- ▶ ROUGE-1 F: 0.517859 P: 0.532602 R: 0.518565
- ▶ ROUGE-2 F: 0.283772 P: 0.291695 R: 0.284565
- ▶ ROUGE-L F: 0.517599 P: 0.538643 R: 0.510171

METEOR 指标

- ▶ METEOR: 0.508303

注意力可视化

