



第7章：序列到序列模型

课件制作：曲彦儒、朱耀明、屠可伟
讲解人：曲彦儒、朱耀明

定义

序列到序列模型

- ▣ 序列到序列 (sequence to sequence, seq2seq) 是指输入和输出各是一个序列 (例如一句话) 的任务
- ▣ 序列到序列有非常多的重要应用:
 - 机器翻译 (machine translation)
 - 问答 (question answering)
 - 对话 (dialog)
 - etc.

应用

seq2seq

- 一般的sequence-to-sequence模型由两部分组成，一个为编码器(encoder)，一个为解码器(decoder)
- 以中文-英文翻译为例
 - 编码器接受中文输入，将其编码为一个向量
 - 解码器接受编码器生成的向量，然后进行英文生成的任务

7.1 基于神经网络的序列到序列模型

- seq2seq模型架构：编码器+解码器
- 循环神经网络：
 - 编码器依次接收源序列的词，最后一步的隐状态作为解码器的初始状态
 - 解码器从起始符（如<sos>）开始，解码出终止符（如<eos>）结束
- 注意力机制：
 - 在解码的每一步，解码器隐状态作为查询，将编码器计算的源序列隐状态作为键和值，计算输出向量
 - 注意力机制使得解码器能够直接“看到”源序列

7.1 基于神经网络的序列到序列模型

□ Transformer模型：

- 编码器与Transformer语言模型几乎相同，除了以下区别：
 - 不需要像语言模型那样，每一步只能看到前置序列，而是看到完整的句子，因此去除了注意力掩码
 - 去掉了用于输出的线性分类器
- 解码器同样与Transformer语言模型几乎相同，除了：
 - 增加了交叉注意力模块，以便在解码时引入编码器所计算的源序列信息

7.2 学习

seq2seq

- seq2seq的训练方法和普通的语言模型**基本一致**，即**使用MLE进行训练**，唯一的区别在于，它每一步会同时用到源序列和目标序列
- 训练seq2seq模型的常用方法为教师强制（teacher forcing）：
 - 使用真实的目标序列作为解码器的输入
 - 会导致曝光偏差（exposure bias）

7.3 解码

seq2seq

□ 贪心解码:

- 每一步取概率最大的词
- 存在错误累加问题

□ 束搜索解码:

- 每一步保留 k 个最佳的候选结果， k 被称为束宽
- 对于上一步的 k 个候选结果中的每一个，生成 k 个候选结果
- 在 k^2 个新的候选结果中，选取最好的 k 个保留

总结

□ 序列到序列模型

- 编码器与解码器
- 循环神经网络
- 注意力机制
- Transformer模型
- 学习
 - 教师强制
- 解码
 - 贪心解码
 - 束搜索解码