

第四天-聊天机器人的模型

- seq2seq模型
 - 是一个Encoder-Decoder结构的网络，它是输入是一个序列，输出也是一个序列
 - 编码器：Encoder中将一个可变长度的信息序列变为固定长度的向量表达，解码器：Decoder将这个固定长度的向量变成可变长度的目标的信号序列
 - 这个结构最重要的地方在于输入序列和输出序列的长度是可变的
 - 可以用于翻译，聊天机器人，句法分析，文本摘要等
-
- 注意力机制
 - 是序列到序列模型中用于注意编码器状态的最常用方法，同时还可用于回顾序列模型的过去状态
 - 不仅能够来处理编码器或者前面隐藏层，它同样还能用来获得其他特征的分布，例如阅读理解任务中作为文本的词向量
 - 为什么要引入注意力机制？
 - 减少处理高维输入数据的计算负担，通过结构化的选取输入的子集，降低数据维度
 - 让任务处理系统更专注于找到输入数据中显著的与输出相关的有用信息，从而提高输出的质量
 - Attention模型的最终目的是帮助类似编码解码器这样的框架，更好的挖掘多种内容的模态之间的相互关系，从而更好的表示这样写信息，克服其无法解释从而很难设计的缺陷
 -
-
- 聊天机器人根据对话的产生方式
 - 基于检索的模型
 - 数据库 - 保存的都是问答对，基于现有的，规定好的数据集
 - 优点：有明确的问答对数据集，使用语句匹配的形式查找答案，答案相对固定，且很少出现语法错误，不会出现新的语句
 - 缺点：重复的语句较多，数据要求较大
 - 基于生成式模型
 - 优点：
 - 不依赖预先设定的问答库
 - 通常基于机器翻译技术
 - 缺点：
 - 答非所问
 - 需要大量的语料进行训练
 - 基于混合模型 - 检索模式和生成模型混合
 - 兼具检索模式和生成模式
 - 目前最常用的解决方法
 - 检索模式产生候选数据集

- 生成模式产生最终答案

-
-
-
-
-
-
-
-
-
-
-
-
-