- NLP方向总结-文本摘要
  - 论文
    - Learning to Encode Text as Human-Readable Summaries using Generative Adversarial Networks
      - 对抗生成式文本摘要
        - 特点：无监督，组成部分：一个生成器、一个判别器和一个重构器
        - 结构：图1
          - 生成器G
            - 生成器和重构器组成AE自编码结构
              - 是seq2seq混合指针-生成器网络,可以决定通过指针从编码器输入文本中复制单词或从单词库中生成
            - 生成器和判别器组成生成对抗网络GAN
            - 生成器负责把长文本编码成短的文本
          - 重构器R
            - 重构器负责把生成器的短文本还原成长文本
          - 判别器D
            - 判别器负责判断短文本是否满足人类可读的要求
        - 模型损失
          - 重构器损失： 源文档x和重构器的输出x^之间的交叉熵损失，输出尽可能地接近原始文本x
            - 离散序列是不可微，使用强化学习构建奖励部分
          - 判别器损失：
            - 方法1: Wasserstein GAN
              - 将生成器的输出概率分布直接反馈给判别器
            - 方法2:对抗性的REINFORCE
              - 将采样的离散词序列送入判别器，并从判别器中评估序列的质量，作为对生成器的奖励信号。
    - Text Summarization with Pretrained Encoders
      - 文本摘要
        - 定义
          - 提取长文本的中心思想
        - 分类
          - 抽取式
            - 抽取文件中最重要的句子来创建总结
            - 二分类问题
              - 一个分类器预测哪些句子应被选为摘要。
          - 抽象式
            - 生成文本的摘要
            - 生成问题： seq2seq
        - 结构
          - 编码器
            - BERTSUM
              - 编码器输入类似:CLS,Sentence1,SEP,CLS,Sentence2,SEP,CLS,Sentence3,SEP
          - 抽取式（BERTSUMEXTABS）
            - BERTSUM编码器+2层Transformer+Sigmoid分类器
            - 二分类交叉熵损失
          - 抽象式（BERTSUMABS）
            - BERTSUM编码器
              - 分别优化，优化器1，Adam
                - lr = 2e-3, and warmup = 20, 000
            - 解码器：随机初始化的6层Transformer
              - 分别优化，优化器2，Adam
                - lr= 0.1, and warmup = D10, 000
          - 实验数据集（英文）
            - CNN/DailyMail新闻摘要数据集
            - 纽约时报注释语料库
            - XSum
          - 评估指标
            - ROUGE-N:系统和参考摘要之间N-grams的重叠。
            - ROUGE-1:指的是系统和参考摘要之间的unigram（每个字）的重叠情况。 本文采用
            - ROUGE-2:指的是系统和参考摘要之间的bigram重叠。 本文采用
            - ROUGE-L:基于最长共同子序列（LCS）的统计。最长共同子序列问题自然考虑到了句子层面的结构相似性，并自动识别序列中最长的共同出现的n-grams。 本文采用
            - ROUGE-W:基于加权的LCS统计，倾向于连续的LCSs。
            - ROUGE-S:基于 Skip-bigram 的共现统计。Skip-bigram是指任何一对词在其句子中的序列。
            - ROUGE-SU:基于 Skip-bigram 和 unigram 的共现统计。