

Homework-week7

前三题必做

1. MATCHSUM里面pearl-summary是什么？为什么要找到pearl-summary？

- We define the pearl-summary to be the summary that has a lower sentence-level score but a higher summary-level score. 摘要集得分高但是句子集得分低的摘要
- 因为论文观测的大部分数据集的最佳摘要并不是由句子级摘要得分高的摘要结果组成，很大比例是珍珠摘要组成，珍珠摘要在选择最佳摘要里是一个重要的影响因素
 - we can observe that for all datasets, most of the best-summaries are not made up of the highest-scoring sentences. Specifically, for CNN/DM, only 18.9% of best-summaries are not pearl-summary, indicating sentence-level extractors will easily fall into a local optimization, missing better candidate summaries.
 - In conclusion, the proportion of the pearlsummaries in all the best-summaries is a property to characterize a dataset, which will affect our choices of summarization extractors.

参考论文《Extractive Summarization as Text Matching》

2. 知识蒸馏里参数T（temperature）的意义？

- 参数T是为了让模型产出合适的概率分布结果
- 因为大模型的Softmax输出概率里面包含丰富的信息，不是真值的那些类对应的概率和真值那一类的概率的相对大小关系也包含了很多大模型学习到的信息 所以需要参数T来调整输出的概率分布，而不是单最大化真值那一类的概率

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$$

- 如果T接近于0，则最大的值会越近1，其它值会接近0，就会丢失一些信息。
- 如果T越大，则输出的结果的分布越平缓，相当于平滑的一个作用，起到保留相似信息的作用。
- Our more general solution, called “distillation”, is to raise the temperature of the final softmax until the cumbersome model produces a suitably soft set of targets. We then use the same high temperature when training the small model to match these soft targets.

3. TAPT（任务自适应预训练）是如何操作的？

任务自适应预训练(Task-Adaptive Pretraining , TAPT), 在做任务相关的数据集, 先把训练数据都拿过作为无监督的没有标签的数据继续进行预训练(在原来停止的预训练的地方训练一会), 然后再对特定任务进行finetune。

附加思考题（可做可不做）：

从模型优化的角度, 在推理阶段, 如何更改MATCHSUM的孪生网络结构？

在推理时候, 原先需要原文输入到网络进行推理, 再候选数据输入到网络进行推理, 网络的参数量*2 优化: 用一个网络, 将原文和候选数据*batch_size 一起输入到一个网络进行推理