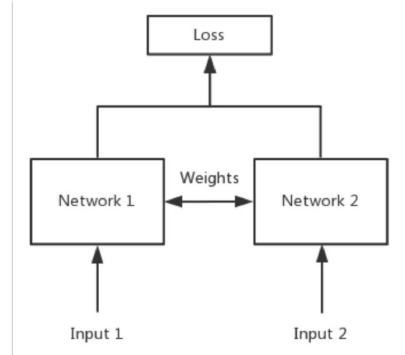
孪生网络和优化策略心得

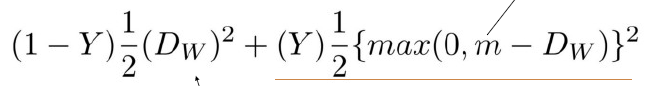
孪生网络

孪生网络，顾名思义则是同一个网络，两个不同的输入，它主要是用来衡量两个输入的相似程度，寻找他们的相同点。它可以用来处理文本匹配，文本搜索，QA聊天等任务。



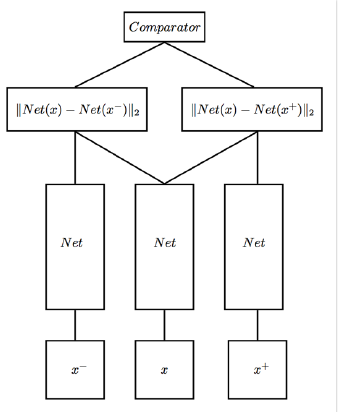
网络结构

损失函数



其中Y=1，表示两个输入相同，Y=0，则表示不同，Dw表示两个输入的距离，可以用欧式距离，余弦距离等等，欧式距离更适合句子级别、段落级别的文本相似度，余弦则是词汇级别

Triple Siamese



网络结构

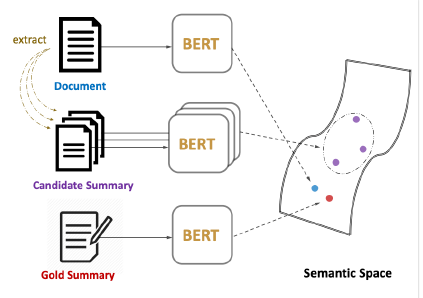
顾名思义，则是三胞胎它有三个输入，Anchor,Positive,Negative。它设计的思想是让正样本离Anchor尽可能的近，负样本则是尽可能的远。所以它的损失函数则是

1623568829(1)

其中，f(A)-f(P)表示Anchor与正样本的距离，f(A)-f(N)表示Anchor与负样本的距离.

MatchSum

孪生网络在文本摘要领域的一个应用。



它的设计思想是将Candidate Summary 投射到语义空间里面，它里面比较好的点跟Document比较近。

Sentence-Level Score表示的是候选摘要中每个句子跟gold summary 的分数。

Summary-Level Score表示的是候选摘要作为一个整体跟gold summary的分数。

首先，提取候选摘要句：基于句⼦级对句⼦打分，选择m 个候选句（采⽤Bertsum的⽅式提取句⼦级摘要），然后从个句⼦中选择 (n可以有多个值)，最终选择了n 个摘要级组合.

然后，基于语义匹配选择最佳摘要: 基于提取出摘要级组合，与原⽂本计算相似度。本⽂考虑了两个loss来finetune，⼀个是基于候选摘要与原⽂档的相似度，其⽬标函数为

1623569327(1)

另⼀个考虑候选摘要之间的差异性，即基于margin loss的思想，认为得分靠前的与得分排后的有较⼤的差⺠，其损失函数可表示为

1623569348(1)

整体的损失函数

L=L1+L2

其中f(D,C)表示的是文本与候选摘要之间的相似度，1623569582(1)表示的是文本跟Gold Summary之间的相似度。

预训练技巧

在我们训练网络时候，存在着一些不更改模型的前提下，存在和一些方法，取得更好的效果。

DAPT

首先将我们的预训练模型，分别使用4个领域的数据做第二阶段的预训练，生成4个领域自适应的语言模型，然后将自适应的语言模型去做相应任务，这是希望预训练模型能学到任务对应领域的一些特性。如果预训练模型之前没有经过任务对应领域数据越多，这种效果越好。

TAPT

首先将我们的预训练模型，然后将各个任务本身的训练数据当作⽆标签数据来进⾏第⼆阶段的预训练，得到任务⾃适应的预训练语⾔模型，然后将自适应的语言模型去做相应任务。

与DAPT相⽐，TAPT使⽤的是更⼩的，但与⽬标任务更相关的预训练数据集。为了TAPT效果，我们能有与⽬标任务数据分布相同的更⼤的⽆标签数据集。可以有

1）⼈⼯提供较⼤型⽆标签数据集；

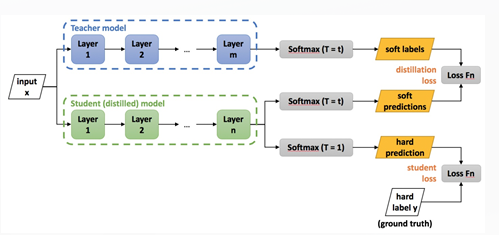
2）使⽤算法⾃动在领域数据中挑选出与任务数据分布相同的数据集

DAPT+TAPT一起的效果更佳。

模型压缩

模型蒸馏

它是将一个较大的、效果较好的ensemble模型压缩成一个轻量级的模型，同时解决labeled data不足问题。先用过ensemble模型对大量的unlabeled data 打标签得到label，然后对这部分样本训练一个轻量的模型。它的目标是我们不仅有正确的目标，还希望Soft target分布尽可能相同，它包括两个模型teacher model 和student model



它要求student model在训练时，除了要使得其输出与true label尽量拟合外，也要求最后softmax之前的logits和teacher network中对应的logits，在一个temperature参数T的变换下，能尽量接近。



其中Zs是student,Zt则是老师

量化

降低模型参数精度，从而降低内存占用，提升速度。

提升推断速度

1使用tensorRT加载onnx

将Transformer用Reformer代替。