关于这一项目的心得

摘要，顾名思义，是从大量的文本数据中提取关键信息，以便用户抓住核心。提取摘要有两种方式，一种是生成式，一种是提取式。提取式也就是从原文中将文章中一些重要的句子提取出来。它的方法有Text Rank等。

生成式，则是赋予了我们大量的创造性，我们可以使用我们想到的语言去组织描述它。这一类主要的模型开始从seq2seq。seq2seq是一个Encoder2Decoder结构的网络。Encoder的主要作用就是将输入信息提取成一个context。然后将context传入到Decoder中。但是decoder中从encoder中获取的信息不可能是均衡的。当我们输入信息过长，Encoder2Decoder会丢掉一些信息。所以我们需要有侧重点，这就引入了attention的机制。Attention它的思想是一种权重分配的思想。由于attention中权重的计算与位置无关，这显然是不合理的，这就需要位置编码。当我们decoder生成文本时，是将前一个生成信息作为下一个的传入，如果作为下一个输入是错误的信息，则整个训练效果变得很差。这就引入了NLG的思想，这里面，我们可以选择greed decoding、Beam Decoding方法，也可以选择合理的采样方法，这包括Top-k和Top-p。Encoder2Decoder不能并行进行，它对长距离的依赖效果不佳，这就引入了transformer结构。它里面包括了self-attention和Multi-Head attention以及Postional Encoding。Bert是一种双向的transformer预训练模型，它包含了预训练的思想和以及两个预训练任务，它可以解决一词多义的问题，我们可以用bert模型来处理摘要、QA等任务。也可以用来做提取词向量的一种方法，后面接其他的层。它也可以用做BertSumExt和BertSumAbs等任务。但是它主要是用来做抽取式的任务。但是有时候我们在做任务时候既需要保留一些重要信息例如品牌，标签等，同时又需要能给与它生成的能力，我们可以采用PGN模型。当我们用bert模型比较大的时候，我们可以采用蒸馏的思想采用distillBert。它既减小了模型的参数，同时也能保留bert的大部分能力。当我们需要更好的训练的效果，我们需要进行参数优化，其中学习率占重要部分，我们需要动态设置学习率，前期大后期变小，还需要设计⼀个会持续变换超参数实施优化主要以log域的尝试。我们也可以尝试其他模型类似Albert、Roberta等其他模型，他们从一定程度上各种这样那样的方式解决了bert中存在的问题（例如训练长度限制，与下游模型不匹配问题等）。我们在不更改模型基础的前提下还可以采用DAPT和TAPT去取得更好的效果。最后模型的评估准则主要采用roug-n,roug-l。