论文笔记《Decomposable neural paraphrase generation》

论文主体内容是提出了一个从phrase level(包括word和phrase级别)、sentence level级别对句子进行复述。

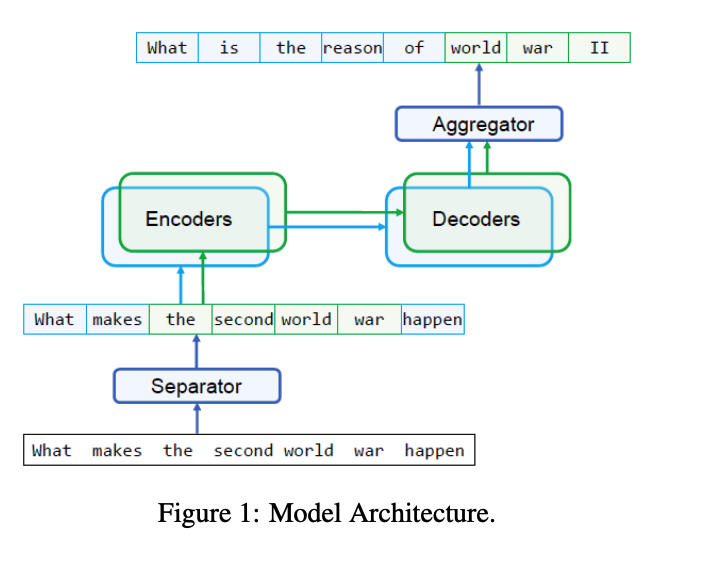
Contribution:

1. 提出了一个seq2seq模型【DNPG模型】来从不同的粒度（短语级别、句子级别）对句子进行复述。
2. 验证了提出的模型在生成复述句子时更易理解和控制
3. 基于提出的模型，提出了一种无监督简单有效的方式在不同的领域试用【不同的语料】。

Model Overview

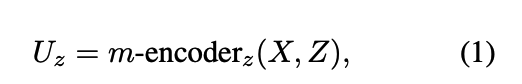
DNPG模型主要包括四个部分：separator、multi-granularity encoders and decoders、aggregator。

大体图像如下：

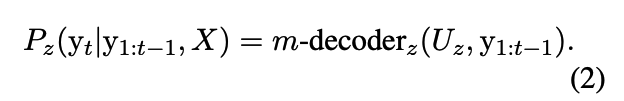


输入会有两部分，一个是句长为L的句子的词序列, separator需要首先确定每个词的粒度【phrase/sentence level】,对应生成.

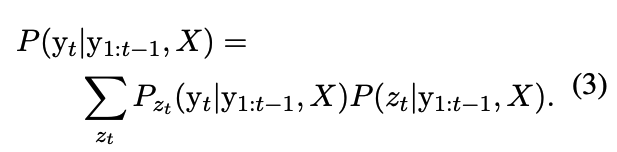
然后将X和Z一起输入到encoder部分：

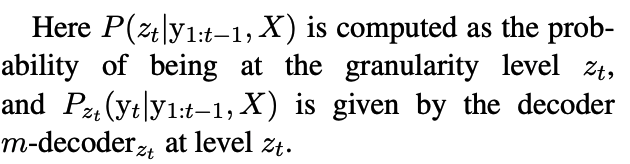


在decoder部分，分别计算生成词的概率大小：



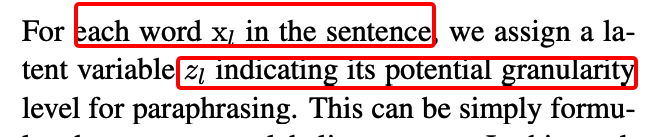
最后在aggregator部分，将所有的decoder部分联合在一起来预测最后的生成词。

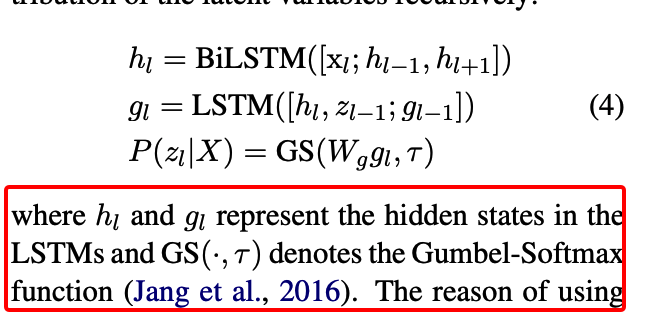




下面对模型的四大部分来单一进行介绍：

1、**Separator**

****

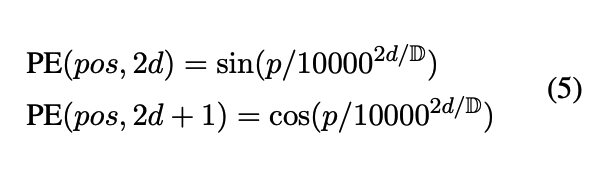
****

**查一下什么是Gumbel-Softmax。**

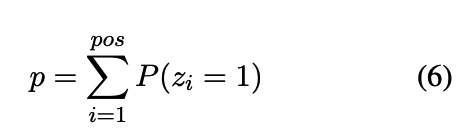
**2、3、Multi-granularity encoder and decoder**

**采用的是transformer的结构，其中phrase-level 对应的是和，sentence-level对应的是和.**

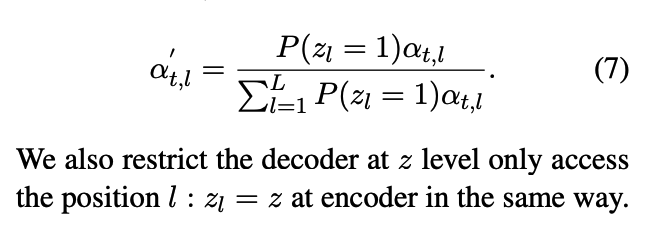
**都加上了positional encoding, phrase level的pos encode如下：**

****

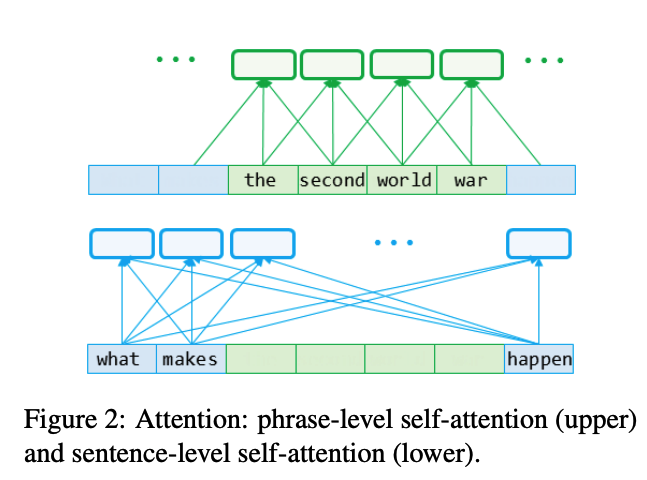
**Sentence-level的为了体现句子的长度，计算pos encode时，需要加上前面句子词的pos encode,公式如下：**

****

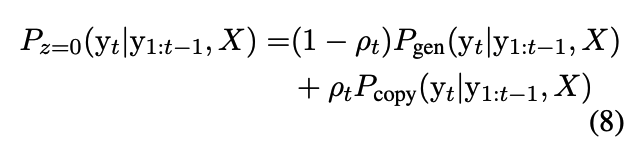
**在计算attention的时候，phrase-level只计算他们相连的三个词（n=3）,sentence-level允许覆盖整个句子中的label为句子级别的词，最后有个softmax对它们进行归一化：**

****

**Attention的具体示意图如下：**

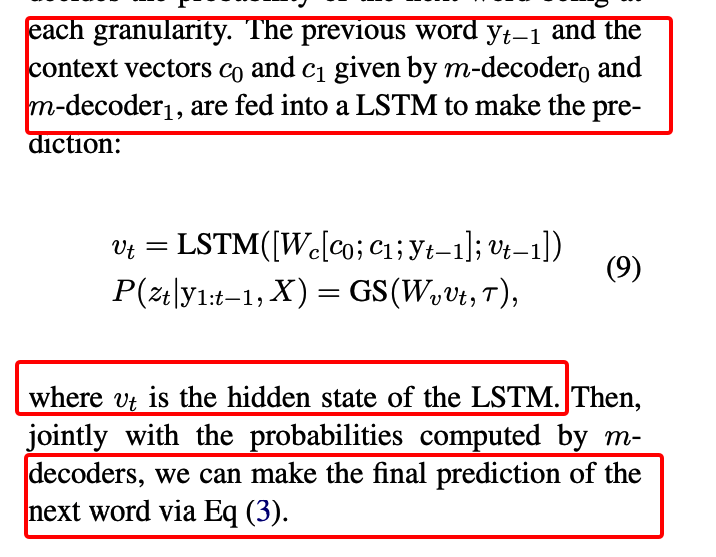
****

**最后decoder部分生成预测词的概率如下：**

****

**4、Aggregator**

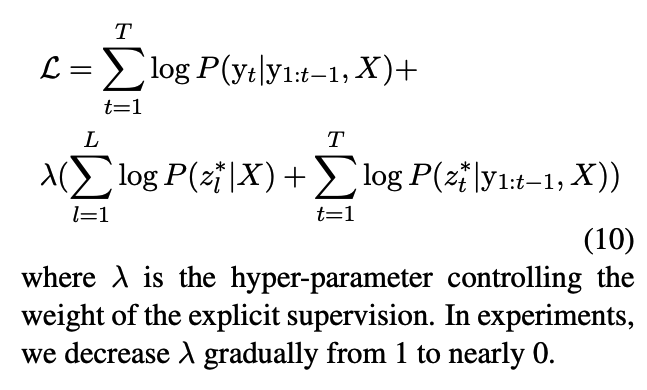
**把两个级别的encoder部分合起来，输入lstm得到一个编码结果，也就是说aggregator从所有的粒度考虑，然后决定生成哪个预测词：**

****

**最后可以用公式3预测最后的生成词。**

**生成损失函数的过程目前还不是很清楚，文中说，首先用word alignment model句子对间产生词的连接，然后将z=0（phrase level）的随机为n=1,2,3，然后剩余的部分就都是z=1的（sentence level）.**

**损失函数如下：**

****

**后续的实验部分介绍了phrase-level、sentence-level的复述的具体case，以及提取的复述模板等。**