阅读笔记Controllable paraphrase generation with a syntactic exemplar

来源：2019 ACL

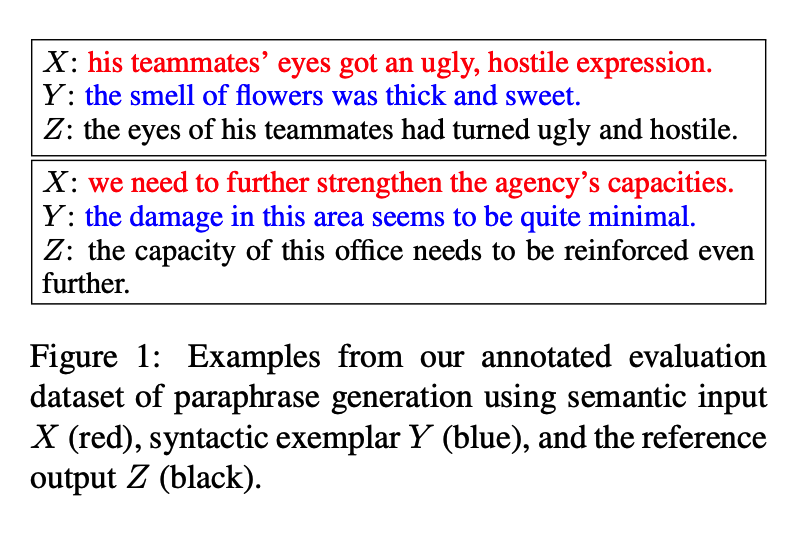
文章主要内容：

对于生成可控的paraphrase进行实验，作者提出了一个生成模型有两个隐变量，一个用来捕获语义信息(semantic)，一个用来捕获语法信息(syntax)。而且还有一个多任务学习，来充分利用paraphrase和word order【这部分目前还没看明白】。

作者创建了一个数据集【只有1300条】，每个数据是个三元组，包括：

* 1）semantic exemplar sentence
* 2）syntactic exemplar sentence
* 3）reference sentence：incorporate the semantics of the semantic exemplar and the syntax of the syntactic exemplar【也就是reference sentence有1中的语义，语法形式和2相同】

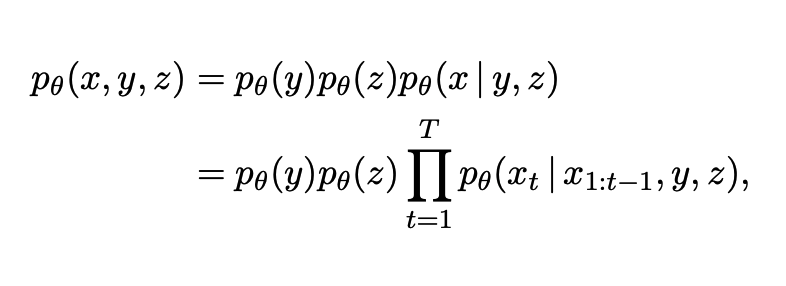
下面是一个三元组的例子：



文中采用了一个SCPN模型【syntactically controlled paraphrase network，后续要看下这个网络】作为基准模型。

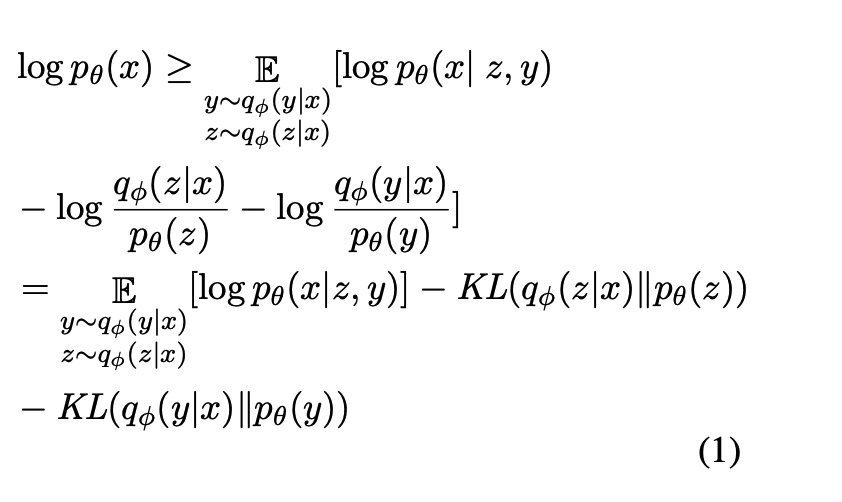
**Method**

实验的目标是给与两个句子x和y，要生成一个句子z，句子z有y的语法形式和x的语义。用数学的方法来表示如下：



作者的实验是沿着上一篇论文A multi-task approach for disentangling syntax and semantics in sentence representations 2019 ACL继续做的，所以有很多的模型细节没有介绍，后续需要再看下这篇论文。

的似然概率有个最低值【在VGVAE中有用到这个公式，后续查一下】：



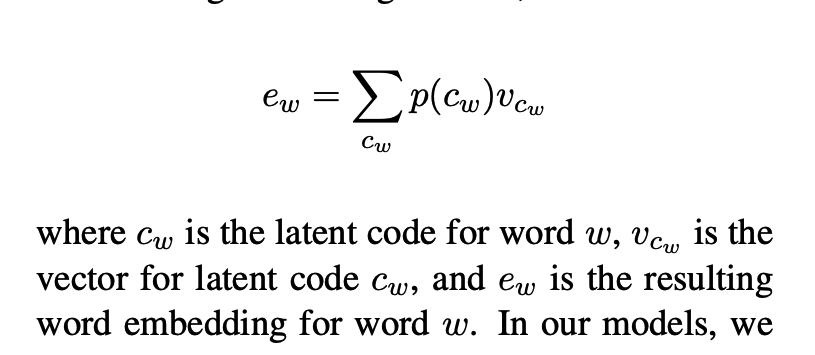
文中假设符合冯米塞斯分布，符合高斯分布。

接下来就是encoder和decoder部分：

* Encoder：用一个三层的前向网络来encode，先用一个双向LSTM，再用一个三层的前向网络来。这两种encode最后都用词的平均来作为最后的结果。
* Decoder：作者这里说在decoder的输入部分都是将语法变量z和前一时刻的word embedding结合起来，输出部分是将语义变量y和隐层变量结合在一起。【这里作者没说用什么做decode，我觉得是LSTM，有可能是采用的别人的模型，就没有详细介绍网络结构，这里可以看下源码】

**Latent coder for syntactic encoder:**

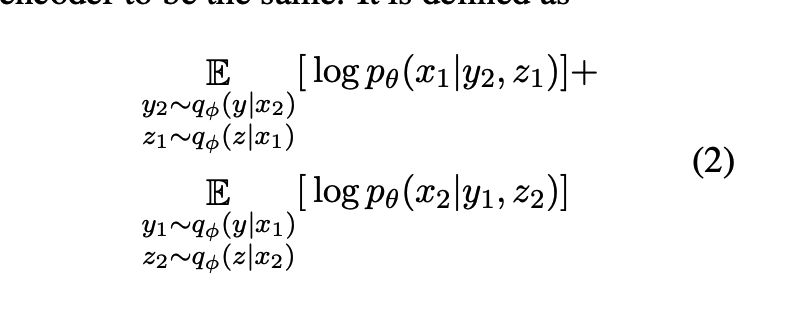
我们希望从syntactic encoder获取的只是syntactic的结构，如果只使用标准的word embedding会导致encoder部分偏向于syntax更注重于具体的词，而我们需要的是syntactic的结构。为了避免这个，作者沿着前人采用的方法，将每个词的隐层代码LC和word embedding结合起来了【这个后续可以继续看一下前人提出这种方法的论文Variational sequential labelers for semi-supervised learning。】



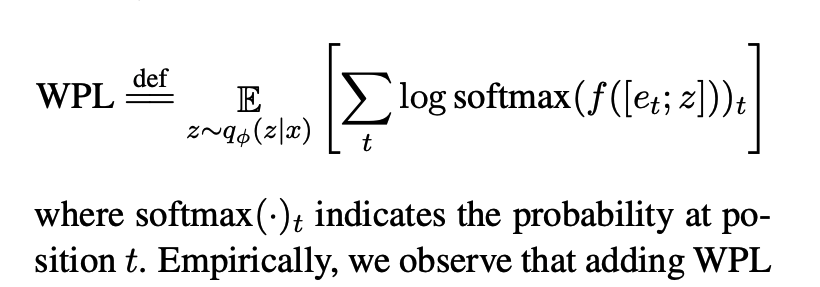
**Multi-Task Learning**

作者设计了两种不同的损失函数，来实现多任务学习：

Paraphrase reconstruction Loss:



Word Position Loss



后续就是一些实验的细节。