论文笔记《unsupervised controllable text formalization》

论文来源： 2019 AAAI

论文代码：https://github.com/ parajain/uctf

论文主要内容：

文中提出了一种非监督的可控的自然语言转换的框架。文中主要实现了以下两个功能：

1. 对于一个给定的文本，将他转换为formal style
2. 通过输入control变量控制文本的formalness大小。

文中的可控，主要指文本的readability level、politeness level等。

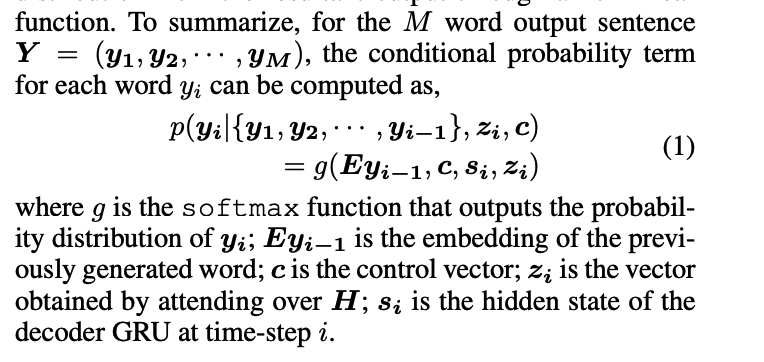
**系统框架**

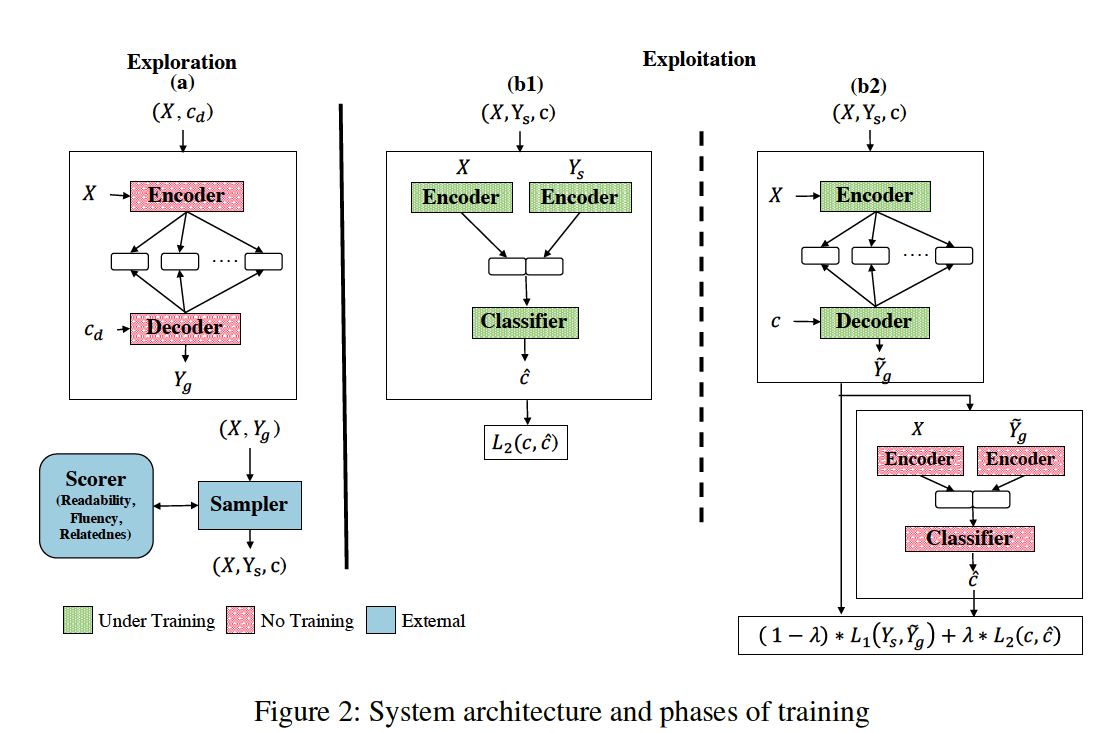
整个框架还是建立在encoder-decoder框架上的，与传统的框架不同的是，文中的结构在decoder部分加入了一个control input。

**Encoder and controllable-decoder**

对于一个给定的输入句子X[由N个词组成]，还有一个d维的control vector c。首先用一个stack GRU对X进行encode，获得隐层表示H。在decoder部分，分为以下三步：

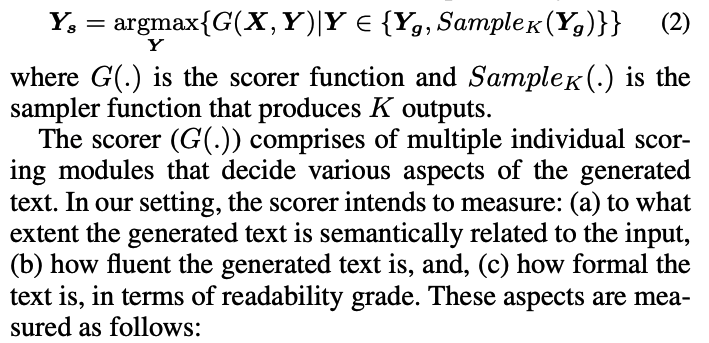
1. Attending：用encoder部分获得的隐层表示来构建context vector
2. Concatenate：将context vector、control vector、current hidden state这三个向量concatenate在一起。
3. 产生一个条件概率分布，计算方式如下：





**Sampler and Scorer**

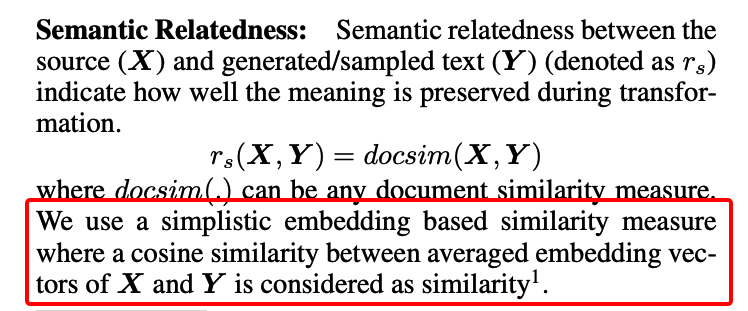
对于一个文本X，decoder部分能产生输出，Sampler部分的功能就是基于产生的K个变种，然后通过计算最大的weighted score【这个weighted score由scorer module产生】选出最好的,公式表达如下：



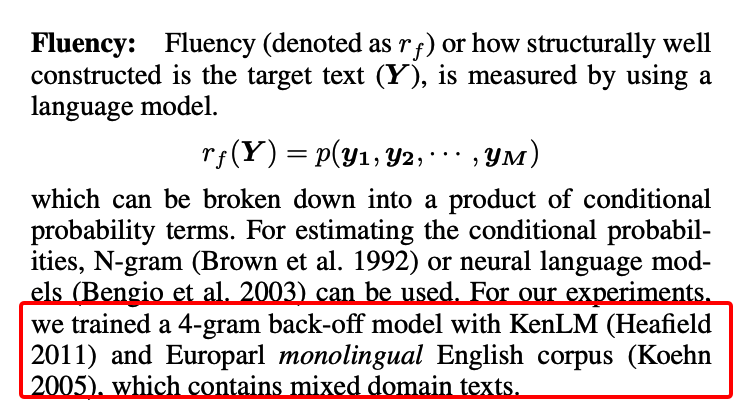
Scorer module主要量化三个方面：

1. Semantic relatedness:生成的文本与输入X在语义上的相关性
2. Fluency：生成的文本的流利度
3. Readability grade：生成的文本的formal程度

**Semantic relatedness**：作者是采用计算X和生成文本Y之间的averaged embedding的相似度。

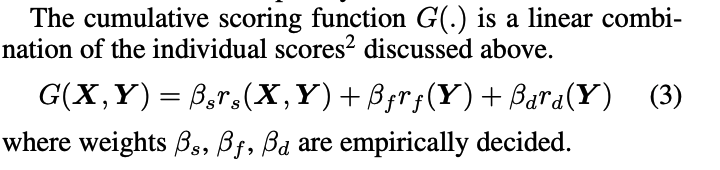


**Fluency：作者基于别的语料训练了一个4-gram的模型，计算产生文本的Y的条件概率：**

****

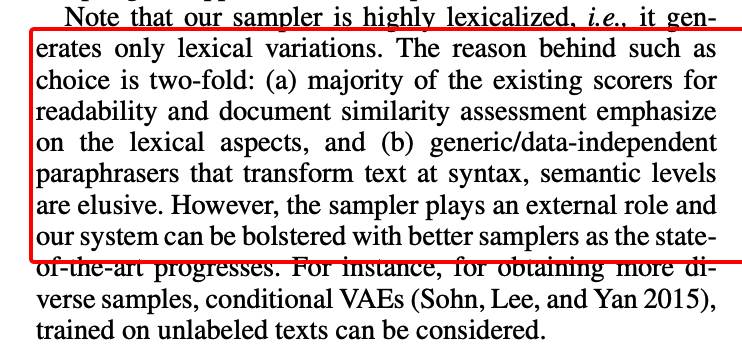
**Readability Grade：一般认为readability score越高，文本的语义复杂度越高，作者选用Flesh-Kincaid readability计算方法。**

**最后的Score model是三个打分的组合：**

****

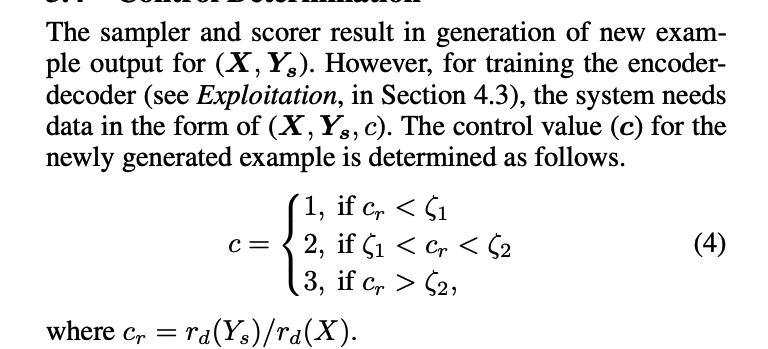
**Sample Strategy**

在生成阶段，对于每个时刻t，除了生成最可能的词外，它的一个同义词也被选择【从同义词表中随机选择】，

****

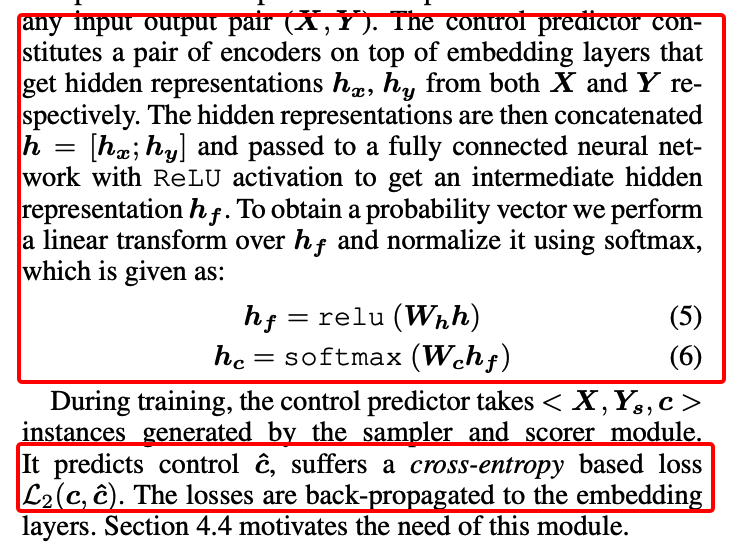
**Control Determination**

**作者通过计算X和Y在readability score上的比值，决定c的大小**

****

**Control Predictor**

**也就是系统图中的b1那一块，预测c值，先分别获取X和Y的隐层表示，然后concatenate在一起，再输入一个全连接层【relu激活函数】，再用一个softmax获取最可能的取值：**

****

**Training Objective and Process**

**这个主要包括三个部分：1）pre-training 2)exploration 3)exploitation，其中pre-training只执行一次，exploration和exploitation在训练中迭代多次。**

**Pre-training**

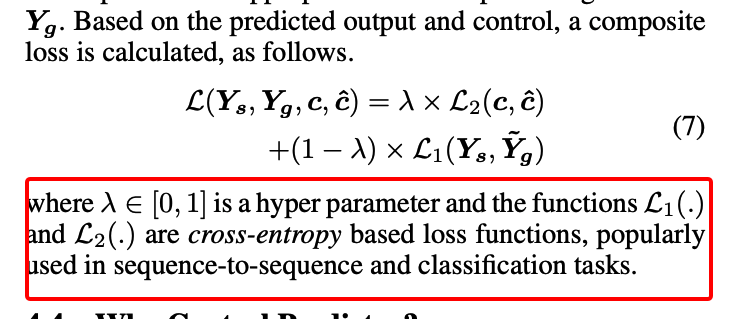
在这一部分，将encoder-decoder框架视为一个autodecoder部分，从X学习，没有control input。

**Exploration**

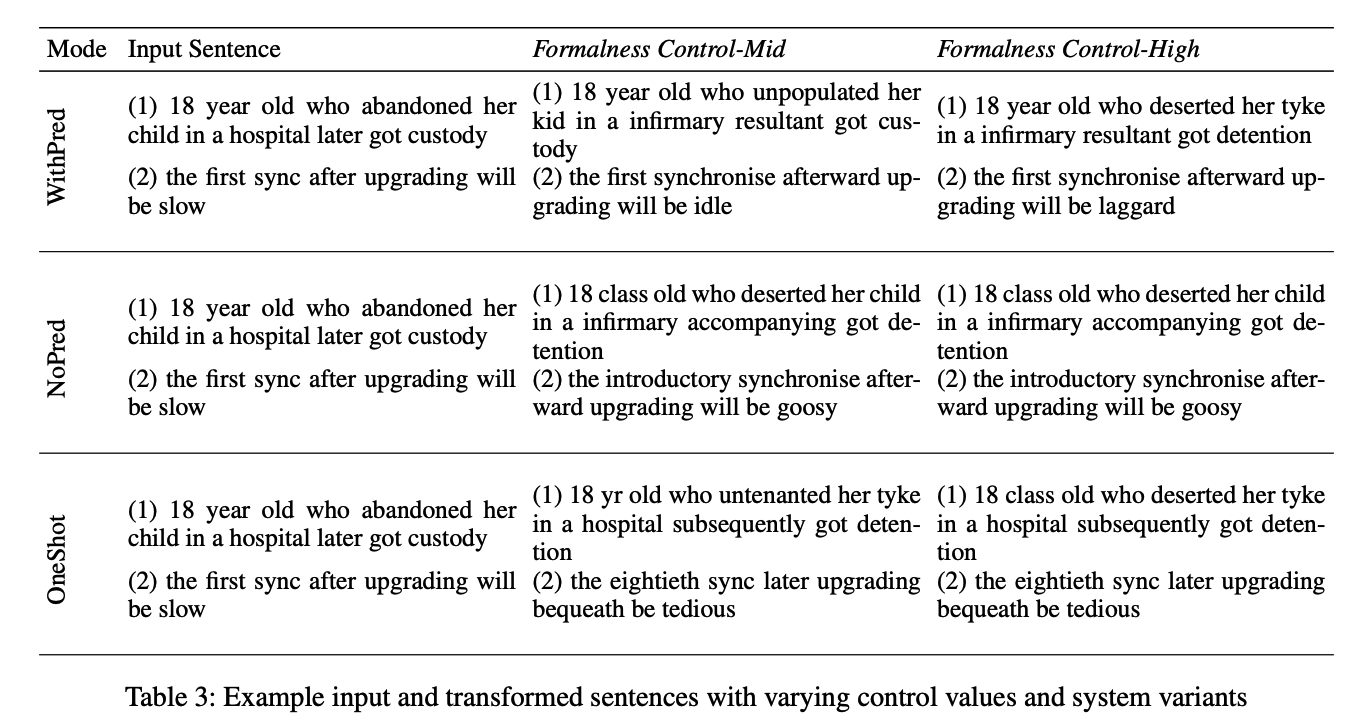
系统框架中的a)部分，具体工作在上面已经介绍了

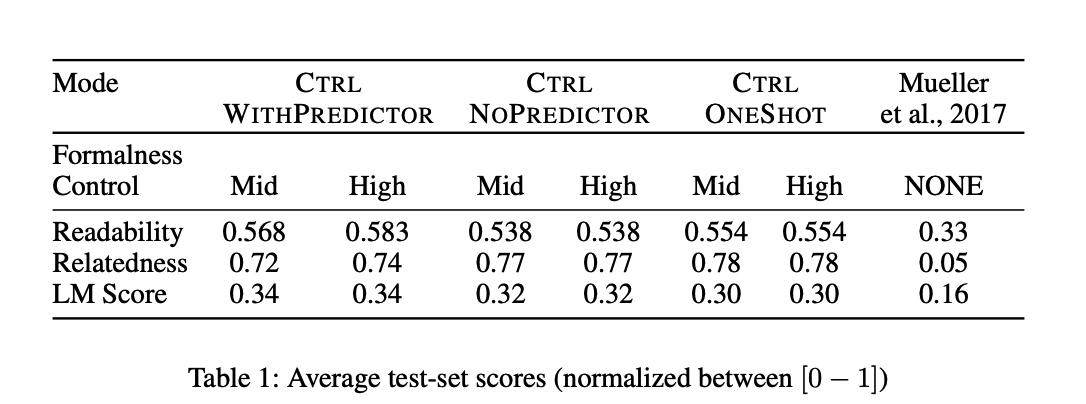
**Exploitation** 这部分主要是系统框架中的b1和b2部分，其中b1的过程在上面介绍过了，b2其实就是a和b1的框架，但是输入变了而已。

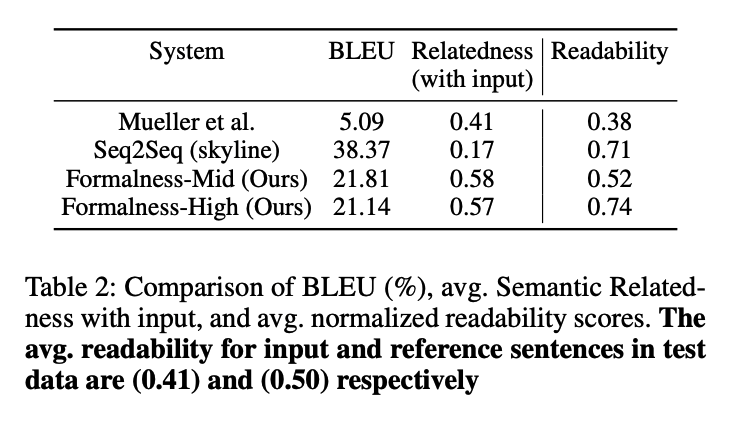
**最后的损失函数为：**

****

**实验结果**

****

****

****