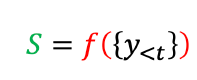
关于NLG的decode相关知识的心得理解

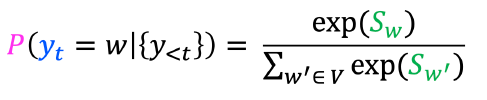
NLG应用

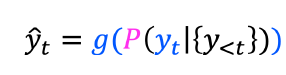
它可以用于翻译，对话系统，文章总结，邮件总结，会议记录，图表生成报告，图画生成描述等等。总之它的应用现在越来越广泛。

它的定义就是所有跟人类文本生成相关的都可以定义为NLG

它的数学定语义是







首先它是根据我们的模型在每一个时间节点上为每一个token生成一个向量分数，然后为它计算概率分布，最后利用这些概率分布和token生成一段话。

Decode方法

全遍历方法

每一步都利用词典V，来进行遍历，然后比较，选择最好，但是如果时间步为T,则它的计算量1621121607(1)就会太大

贪心算法

它则是首先选择生成的概率是最高的单词，然后将这个单词放入下一步作为输入，然后生成的概率分布，然后又选择最高的那个单词，然后又作为下一步输入，与此循环。

它停止的准则是直到最后生成了END或者时间步的限定

问题

1、每一次生成的结果都是固定的

2、一步错步步错

3、局部最优

Beam Decoding

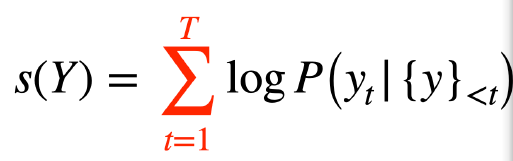
它则是每一步选择概率分布最好的k个，然后用这个k作为输入，与此循环，最后计算分值，看哪个最高。

停止的准则

生成END token，时间步的限定或者至少生成了n个假设。

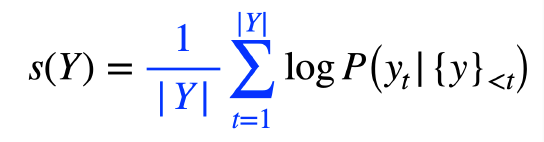
k个选择，如果太小就会Ungrammatical,unnatural,nonsensical,incorret,太大一方面会减少之前问题，但是就会计算量太大，同时生成语句越短，造成BLEU下降太快。所以需要我们探索最终决定。

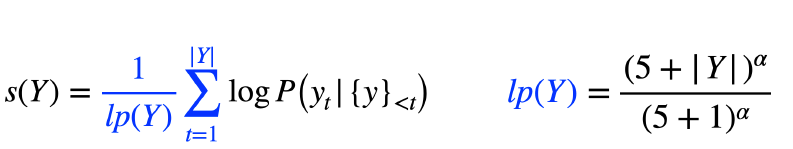
上面说到它会惩罚生成的很长的句子。



从上面公式可以看到，单词越短，分数越好，但是句子越短不利于用户体验，所以需要改正

归一化方法

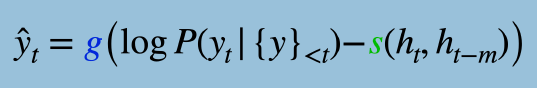




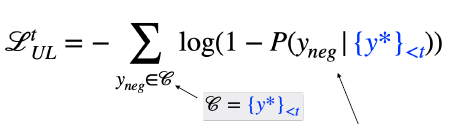
还会造成重复词的问题

解决方法：

1. 手动限制不要重复
2. 增加格外的loss，惩罚相近的hidden state



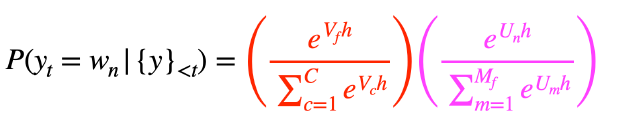
1. 不是似然的loss function



找到一系列token 1621132808(1),它用来减少似然值。

1. F^2 softmax

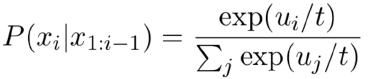
首先先将token频率排序，然后将它分类。我们生成单词时候就可以先频率分类，然后在具体点的那个频率里面在单词分类。分成了两步。



采样策略

在采用beam search，它生成概率很高，并且很稳定，但是人类语言是多样。所以引进了采样策略。

温度参数采样



它放大那些频率高的单词，缩小那些频率低的单词，从而采样概率也就相应变化了。

Top-k采样

选择k个token，在k个里面进行采样。

问题

对于频率差不多的分布时候，top-k就会舍弃差不多的单词，peak的，容易保留很低的单词

Top-p采样（核采样）

按照百分比采样。

它们本质区别在于置信区间的选择。

评估指标

ROUGE-N

ROUGE-N measures the number of matching ‘n-grams’ between our model-generated text and reference。评测n-grams

ROUGE-L

ROUGE-L measures the longest common subsequence (LCS) between our model output and reference。最长子串

ROUGE-S

S for skip-gram