

1.

第一步：

先判断出出错的单词。单词写错的概率。

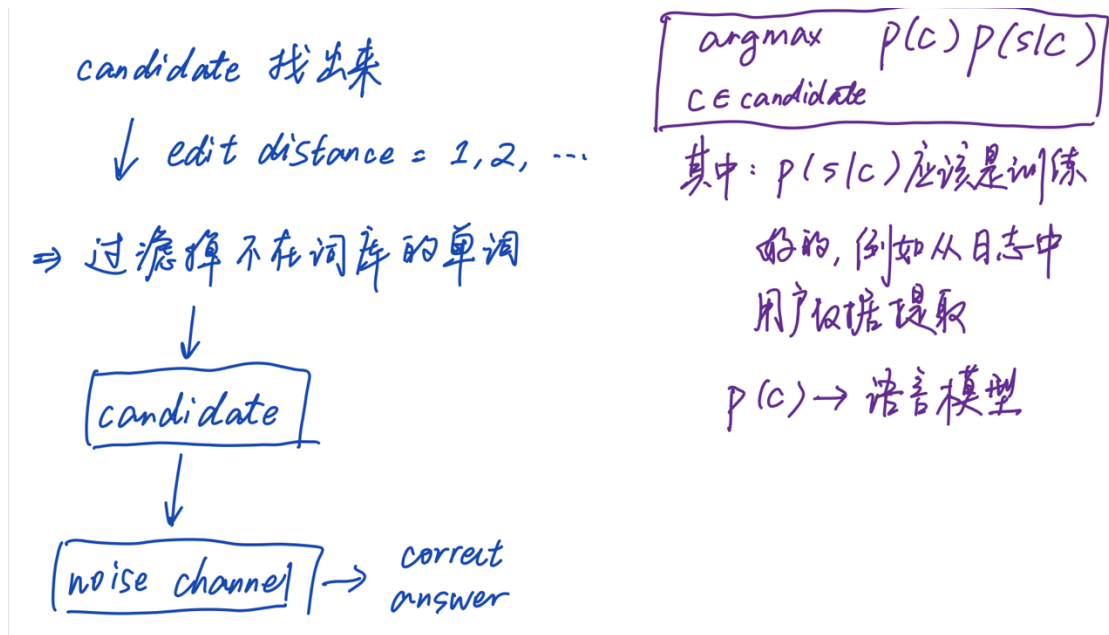
也就是需要考虑语言模型，这个错误的单词在句子中是不通顺的。

候选集合：

围绕当前单词，生成 edit distance 为 1, 2... 的候选单词。

然后将选出来的单词，过滤其中不在词典中的单词。由此找到 candidate words。

再用 noisy channel 选出 correct answer。



2. jupyter

2.1

I like playing soccer.

term-count : (I, like, playing, soccer)

bigram-count : (I like, like playing, playing soccer)

计算出现次数

那么 e.g. $p(\text{like}/I) = \frac{C(I \text{ like})}{C(I)}$

I 也可以看成 bigram : 前面的 <s> 空格时，后面多少次出现 3 I

