模型衍生

一．对于基于Tri-grim模型的单词联想度指标，本文的假设是连续的马尔科夫假设，但是在困难模式中常常会出现不连续的情况，故我们用不连续的马尔科夫假设对本文模型进行进一步优化并给出wordle游戏策略。

基于不连续的马尔科夫假设的Tri-gram模型

构建语料库，我们将单词库中的单词全部划分为

Wi1 wi2 wi3wi4 wi5

W21 w22 w23 w24 w25).\*(n1,n2……n10658)

……

假设单词正确为（r1 r2 r3 r4 r5）

初始值（s1,s2,s3,s4,s5）输入，当我们确定了一个字母位置时，假设为01,02,03,r4,05，,0j表示第j个位置未知，我们从语料库中找到符合01,02,03,a,05形式下的所有单词，并求出每个单词的概率，代入最大概率的单词，若仍未确定第二个字母，选取概率第二大的单词输入，依次操作，直至出现第二个字母确定位置。01,r2 ,03,r4,05,我们从语料库中找到符合01,r2 ,03,r4,05形式下的所有单词，并求出每个单词的概率，代入概率最大的单词，重复操作直至5个单词全部确定完全。

当初始值（s1,s2,s3,s4,s5）输入，确定了2，3，4个字母时，依此操作，直至最优。

游戏策略

由于绝大部分词语含有元音字母a e io u ，我们假定策略是先猜4个元音字母的

Aeidu,确定一个正确字母位置，如果全是灰色，则猜tryst无元音字母。若我们得到1至4个绿色位置，我们就可以采取建立的基于不连续的马尔科夫假设的Tri-gram模型。

二．对于报告结果分布的预测

我们对359个单词的分布做了正态分布检验，其都可以近似为正态分布，但偏度和峰度都有差异。

我们对报告结果分布和偏度和 峰度做了相关性分析



可以发现偏度和峰度在不同程度上与报告结果分布呈相关性。

故我们可以引入偏态分布预测报告结果分布

白板上的文字

描述已自动生成

文本

描述已自动生成

然后我们用参数估计估计参数lamba，得到不同的偏态分布概率密度，帮助我们预测报告结果分布，模型应该会比前文更加准确