本周结合老师的邮件，基于TFDev这个指标进行了各项实验，实验结果整理如下：

1. **上周TFDev实验的修正**

因为上周的实验，在计算TFDev时计算了一元TFDev的值，而老师在邮件说明了，TFDev用来表征一个词汇的独立性，一元词汇的TFDev没有意义。上周实验中在使用归一化TFDev方法时，考虑了一元的值，并且用于了实验。为了修正这个错误，重新做了实验。

对于二元和三元词，采用减去均值方法归一化，即为，而对一元TFDev则指定默认值。最优默认值要通过实验获得。经过实验，一元的默认值最优为：TFDev(1gram)=0.046，此时，平均准确率为82.57%。

1. **VTFDev实验分析**

根据老师在邮件中提及的思想，其实与“Unsupervized Word Segmentation the Case for Mandarin Chinese”中的VBE的思想一致，但是TFDev对于每个词汇只有一个值，不像BE那样既有左分支熵又有右分支熵。并且一个词汇可能含有多个子串（例如ABC的子串就包括AB和BC），具体采用哪个子串去运算还需要考虑。经过对各元数据的分析，我发现，子串中的较小值比较合适。所以此处定义



相应的，各元词汇的VTFDev定义如下：

1. 三元词汇ABC的VTFDev定义为：
2. 对于二元词汇AB，按照定义应为：



但是，因为一元的TFDev无法获得，所以此处需要指定一元词汇的TFDev默认值。即,

所以最终

1. 对于一元词汇A，其VTFDev也无法给出实际的定义，所以也需要定义默认值，即为



最终的VTFdev越小，则一个词的独立性越高（如果觉得这样负相关不好的话，也可以修改减法顺序，则正相关了）。但其中很多词汇也并不满足这一关系。

综上，该方法需要指定两个默认值，才能进行实验，实验中我尝试了很多种组合，最终找到了较优的默认值。

|  |  |  |
| --- | --- | --- |
| 默认TFDev(1gram)default1 | 默认VTFDev(1gram) default2 | 平均准确率 |
| 0.005 | -1 | 0.793629 |
| -0.1 | 0.819583 |
| -0.06 | 0.821741 |
| -0.05 | 0.82264 |
| **-0.04** | **0.822981** |
| -0.03 | 0.82106 |
| -0.02 | 0.82025 |
| -0.01 | 0.816522 |
| -0.004 | 0.811877 |
| -0.002 | 0.808507 |
| -0.0002 | 0.807083 |
| 0 | 0.806766 |
| 0.0001 | 0.802919 |
| 0.0003 | 0.798454 |
| 0.001 | 0.791102 |
| 0.01 | 0.749119 |
| 0.06 | -0.05 | 0.818951 |
| -0.04 | 0.819789 |
| -0.03 | 0.819668 |
| -0.02 | 0.820706 |
| -0.01 | 0.821447 |
| -0.005 | 0.821741 |
| **-0.003** | **0.822182** |
| -0.0015 | 0.821671 |
| 0 | 0.806553 |
| 0.01 | 0.737285 |
| 1 | 0.672095 |
| 4 | 0.663127 |
| 20 | 0.661446 |
| 0.006 | **-0.04** | **0.822962** |
| -0.1 | 0.819702 |
| -0.003 | 0.811877 |
| -0.001 | 0.808507 |
| 0.1 | -1 | 0.793293 |
| -0.04 | 0.813809 |
| -0.01 | 0.818951 |
| 0 | 0.803475 |
| 0 | 0.803475 |
| 1 | 0.668716 |
| 0.001 | -1 | 0.793629 |
| -0.2 | 0.805633 |
| -0.1 | 0.819789 |
| -0.06 | 0.821815 |
| **-0.05** | **0.822693** |
| -0.04 | 0.822541 |
| -0.02 | 0.819493 |
| -0.004 | 0.806766 |
| 0 | 0.783835 |
| 1 | -2 | 0.789684 |
| -1 | 0.793293 |
| -0.04 | 0.793629 |
| -0.01 | 0.793629 |
| 0 | 0.775005 |
| 0.005 | 0.69516 |

其实调节两个参数获得的准确率的上限差不多，大约在82.5%左右。老师上次邮件说的“把各种归一化前后的TFDev的计算结果提取出来，分析取值合理性。”是要观察每次取不同的参数分词后具体哪些地方出错吗？

老师还提到采用除法计算VTFDev,我感觉跟减法是一样的。

1. **NVTFDev实验分析**

其理论类似于第二篇论文中定义的NVBE（**Normalized Variation of Branching** **Entropy**）,根据上面计算得到的VTFDev,定义NVTFDev如下：

，其中是相应元词汇的平均值。根据该公式易得各元词汇的NVTFDev，但类似实验（2）中的问题，需要指定默认值。因为该实验的数据其实是基于VTFDev，而VTFDev计算2gram的指标时，会去指定1gram的默认值。除此之外，本实验也要指定一元词汇NVTFDev（1gram）的默认值，而不能沿用实验2中的数值。

经过实验，尝试各种参数，其实验结果与（2）相差不大

|  |  |  |
| --- | --- | --- |
| 默认TFDev(1gram) | 默认VTFDev(1gram) | 平均准确率 |
| 1 | -1 | 0.793629 |
| -0.2 | 0.805633 |
| -0.1 | 0.819698 |
| -0.05 | 0.821988 |
| -0.047 | 0.822784 |
| **-0.046** | **0.82311** |
| -0.0455 | 0.822878 |
| -0.045 | 0.822609 |
| -0.043 | 0.822889 |
| -0.04 | 0.822339 |
| -0.035 | 0.821043 |
| -0.03 | 0.820537 |
| -0.025 | 0.820517 |
| -0.02 | 0.819085 |
| -0.004 | 8.07E-01 |
| -0.002 | 0.803924 |
| 0 | 0.772708 |
| 1 | 0.680654 |

1. **关于实验结果准确性的度量**

因为前期进行了一些标注工作，所以现在是在未标记语料上进行训练，在标注语料上进行测试，计算一个准确度，计算方法为：分别计算每个句子切分正确的位置占本句所有位置的比例。然后最后求所有句子准确率的平均值。

但这个方法短句切分错误对整体影响较大，例如：句子ABCD一处切错则该句准确率为66.6%，但ABCDEF一处切错准确率则为80%，所以我打算更改下句子准确率的测量，只计算最终切错位置的总数。这也是我在具体分析句子切分错误原因时发现的一个小地方。

1. **关于TFDev具体切错的原因**

老师上次让我详细分析句子切错的原因，我正在思考，有几个地方我还没想好，想好后我再发邮件说明。

1. **结合其他指标提升TFDev分词结果**

考虑如何使用其他的指标对TFDev进行调整，提升TFDev分词的准确性，比如TFDev\*MI^0.5。这个思路我还没有思考，实验目前还没做。

1. **采用其他分词工具重新定义Baseline**

我目前还正在学习怎么使用老师说的另外那几个分词系统。是不是还要基于这几个系统分词后再计算指标，然后跑一下前面的实验？