**参数调整表**

**每个句子与全文的语义相关度计算**

计算每个句子与文章的语义相关度将依据于句子与全文的余弦相似度c1以及与标题的余弦相似度c2的加权。

标题重要性更高，设定C1与C2分别为0.3与0.7

**是否需要knn平滑：**

平滑前：



平滑后：

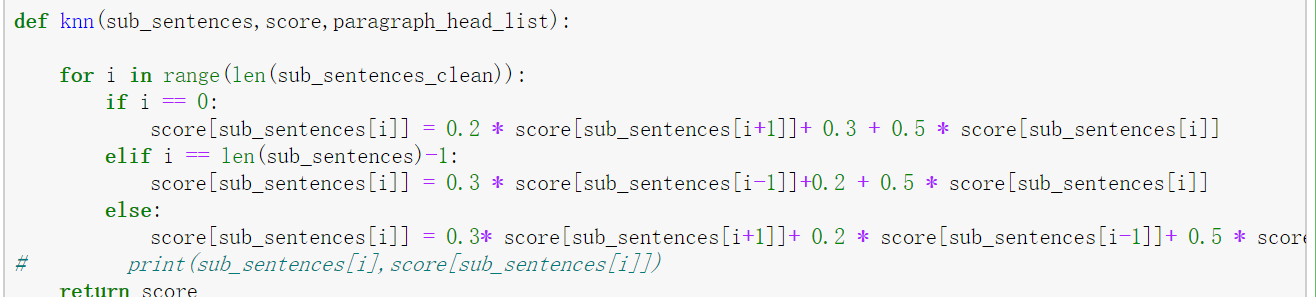


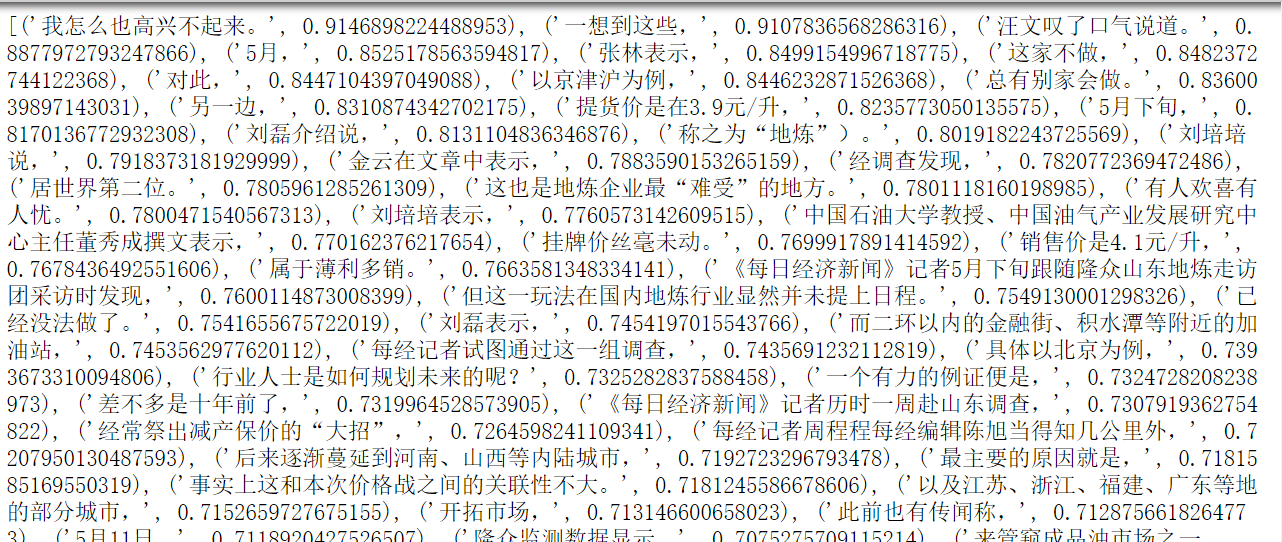
平滑后标准差更小，句子与全文相关度下降。

**Knn平滑：**

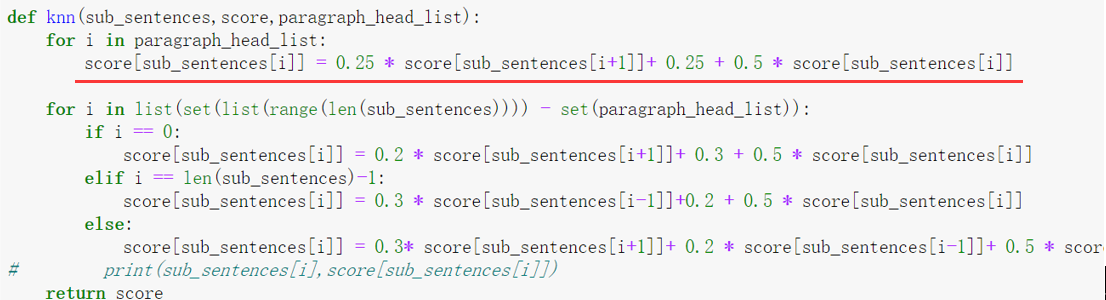
考虑句子的位置：

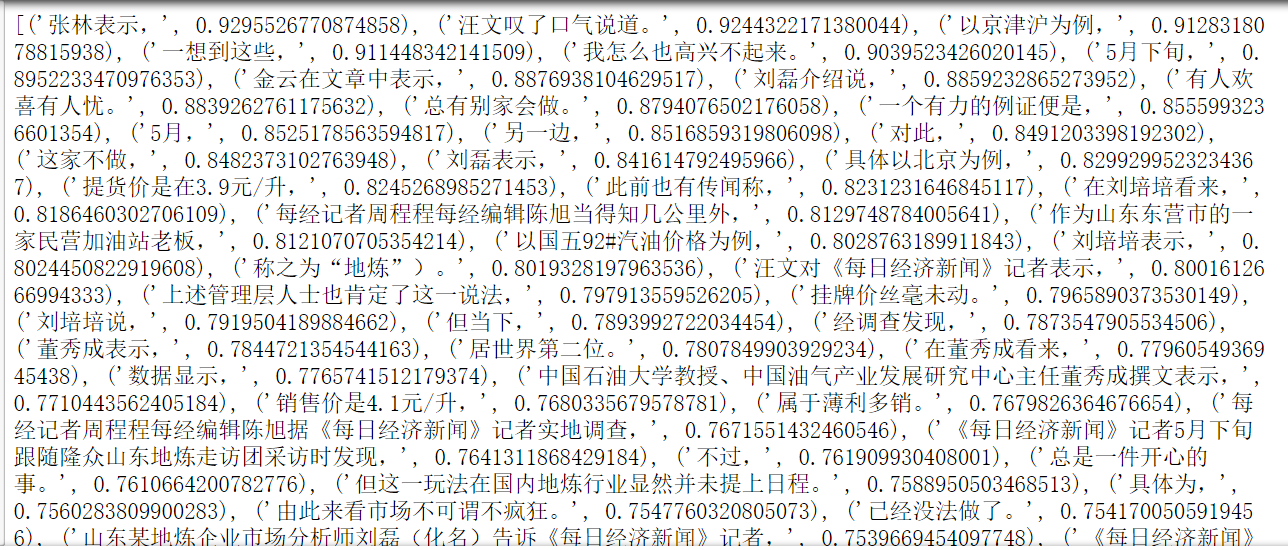
未加强段首句子相关度



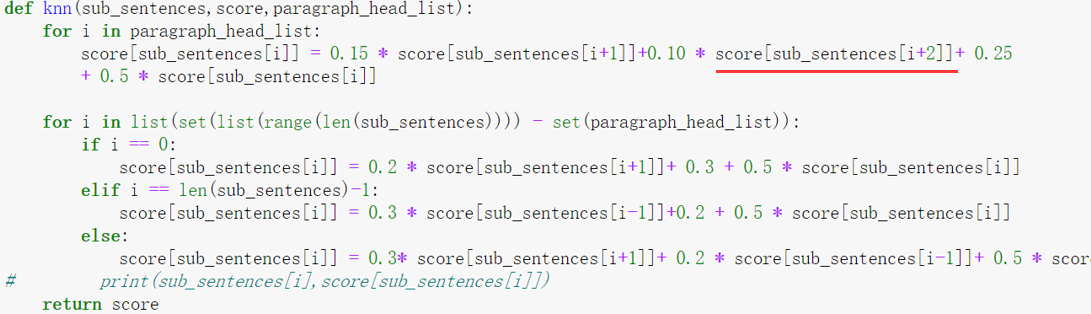


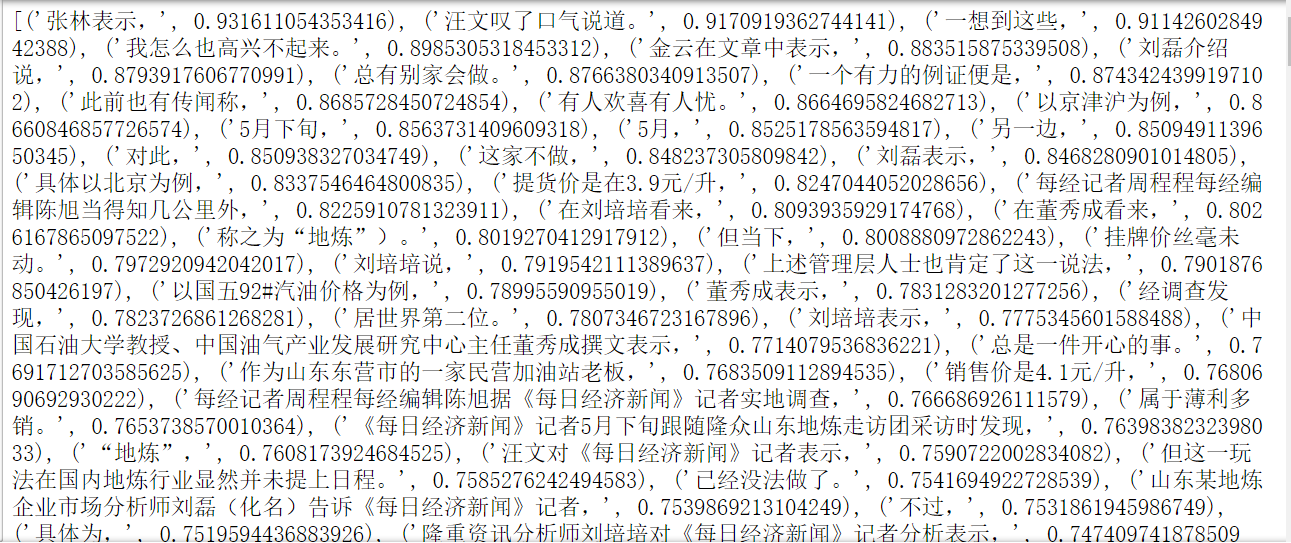
增强段首句子相关度





考虑段首句子与第二相邻的句子有相关性时：





**总结：**

1. 未增强段首句子相关度与增强段首句子相关度相比，段首句子出现的概率较小。

2. 考虑段首句子与第二相邻的句子有相关性时，几个段首句子概率下降。所以摘要生成效果不好。

**句子的全文相关度权重比与全文句子相关度的标准差关系表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 总起句 | 结尾句 | 段首 | 普通句 | 标准差 |
| 1 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.3/0.2/0.5 | 0.135909 |
| 2 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.35/0.25/0.4 | 0.130869 |
| 3 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.4/0.3/0.3 | 0.127545 |
| 4 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | **0.45/0.35/0.2** | 0.125959 |
| 5 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.5/0.4/0.1 | 0.12609 |
| 6 | 0.2/0.3/0.5 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.5/0.35/0.15 | 0.12803 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 7 | 0.2/0.3/0.5 | 0.35/0.25/0.4 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.12596 |
| 8 | 0.2/0.3/0.5 | 0.4/0.3/0.3 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.12597 |
| 9 | 0.2/0.3/0.5 | **0.35/0.2/0.45** | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125958 |
| 10 | 0.2/0.3/0.5 | 0.2/0.2/0.6 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.12596 |
| 11 | 0.3/0.3/0.4 | 0.35/0.2/0.45 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125825 |
| 12 | 0.4/0.3/0.3 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.12571 |
| 13 | 0.5/0.3/0.2 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125642 |
| 14 | 0.5/0.4/0.1 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125663 |
| 15 | 0.6/0.2/0.2 | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125583 |
| 16 | **0.7/0.1/0.2** | 0.3/0.2/0.5 | 0.25/0.25/0.5 | 0.45/0.35/0.2 | 0.125573 |
| 17 | 0.7/0.1/0.2 | 0.3/0.2/0.5 | 0.7/0.25/0.05 | 0.45/0.35/0.2 | 0.120467 |
| 18 | 0.7/0.1/0.2 | 0.3/0.2/0.5 | 0.6/0.25/0.15 | 0.45/0.35/0.2 | 0.11892 |
| 19 | 0.7/0.1/0.2 | 0.3/0.2/0.5 | **0.55/0.25/0.2** | 0.45/0.35/0.2 | 0.118725 |
| 20 | 0.7/0.1/0.2 | 0.3/0.2/0.5 | 0.5/0.25/0.25 | 0.45/0.35/0.2 | 0.11892 |

第19行的标准差最小，所以，我们以第19行参数设置knn平滑函数中的句子的全文相关度权重比。