1. 是否要在新闻材料中筛选出ESG新闻？如何确定ESG新闻的真实影响力，用股票收益率作为ESG新闻影响力的代理变量是否准确？收益率变动中都多少成分是来自ESG新闻的？
2. 筛选出ESG新闻，则忽视了其它新闻对股价的影响，收益率不一定来自ESG新闻的影响；
3. 开始不筛选出ESG新闻，影响力寻址后再筛选出ESG新闻，如何筛选（关键字过滤？）。
4. 在为ESG特征赋权的过程中，对股价越有影响，说明此项特征越重要，在评级时赋予更大权重。那么，是否存在某项特征对股价影响不大，却很能够反映公司的ESG表现呢？换句话说，公司的ESG表现是否与股价有必然对应关系？以这种方法得到的ESG指标，分数越高的公司，股价是否也一定是正向拉升越高的？
5. 利用每一层的Score 评分来确定字词、句、文档的重要性，可视化效果如下：



句子的注意力分数与句子的长短直接相关，同理文章的注意力分数也与文章的长短直接相关，只有同等长度的文字材料其注意力分数才有可比性。

1. 公司每日的新闻集对应了一个已有的影响力大小（日收益率），那么是否可以直接将新闻分为影响力高中低类，在各类中统计指标特征的关键字频数，从而得到该特征的权重。注意力模型除了将影响力权重寻址到字词级别外，字词的影响力权重对构建ESG指标有什么作用？分字、句、文档多层的目的是什么？
2. 如何计算ESG指标的权重。
3. 获得ESG指标权重后，怎样对公司进行评分。

**读期刊：**

1. ESG新闻对证券市场的影响
2. 自然语言处理在证券市场上的应用

**如何将得到词与指标的映射：**提取出所有ESG新闻，进行分词得到词汇字典，用机器学习（CNN分类器）进行词汇分类为每个词注上标签。

**问题：**细分ESG子特征是否影响词分类的准确性？

计算指标权重：

①词score\*句score\*文档score

|  |  |  |
| --- | --- | --- |
| 温室气体排放 |  |  |
|  |  |  |
|  |  |  |
|  |  |  |