上市公司公告文本分析报告

钟东彤 2020200883

**Insight1:词云图**

首先，我通过jieba对爬取的公司公告的文本内容和标题分别进行分词，并将其绘制为词云图如下两图所示：



图1:公告文本内容词云图



图2:公告标题词云图

从图1 中可以看出，近期太钢不锈集团的公司公告围绕着企业日常经营的财务信息、企业分红以及股权激励等展开，其中尤其突出的是公允价值，由于太钢不绣集团为正常参与A股交易的上市公司，因此这里的公允价值应该更偏向交易中的股票报价，说明公司非常关注其自身股价的波动，毕竟这将直接影响公司的估值预期和市场对其的投资情绪。其次，我看到股权激励也较为突出，说明公司目前仍然处于快速发展期，需要大量的股权激励来激励创新以及巩固组织性，这与太钢不绣本身不锈钢行业技术领军者的企业定位是吻合的。此外，我还关注到大量的限售股解禁信息，这与太钢不绣的国企基底与前身不无关系，以前的国企会存有相当部分的成本较低但同股同权的法人股，由于股权分置改革，之前未流通的法人股本也将随时间推移获得在市场自由交易的能力，这在一定程度上活化了公司的资本结构，也在侧面反映出我国证券市场近年来的逐渐放开。

从图2中可以看出，由于是标题与正文的关系，两幅图的主要内容大差不差，但是与图1不同的是，图2更能够从主题层面揭示企业近5个月的公告分布结构。从图2中看到不锈钢作为集团的绝对主营业务的强势业务，占据了绝对的大头。除此之外，日常的股东大会和股份的流转也较为常见，股东大会和董事会的频繁召开也在一定程度上说明近期的人员调度和企业规划转变的频率较高。由于中期报告的时间也包含其中，由此也会带来对外披露年报和财报的需求，因此对外说明会等也占一定比重。

**Insight2:时间序列条形图**

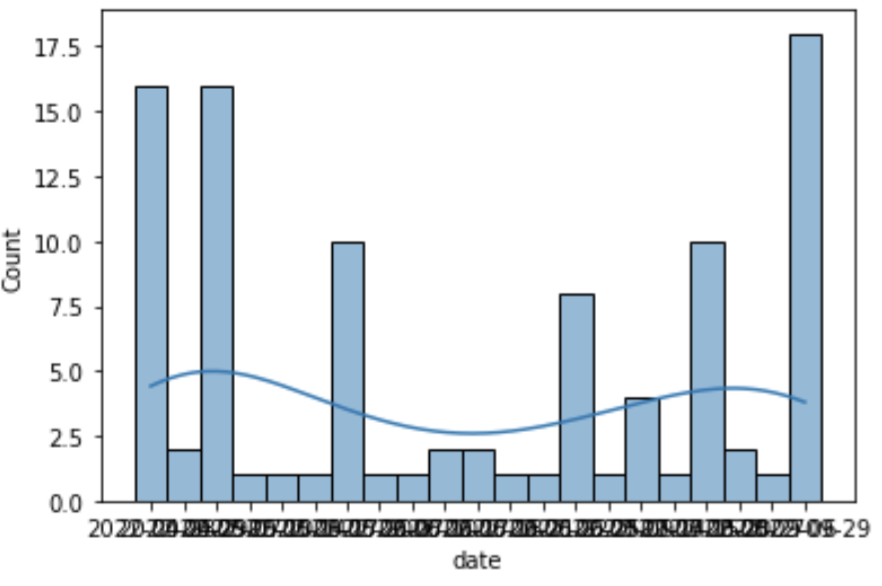
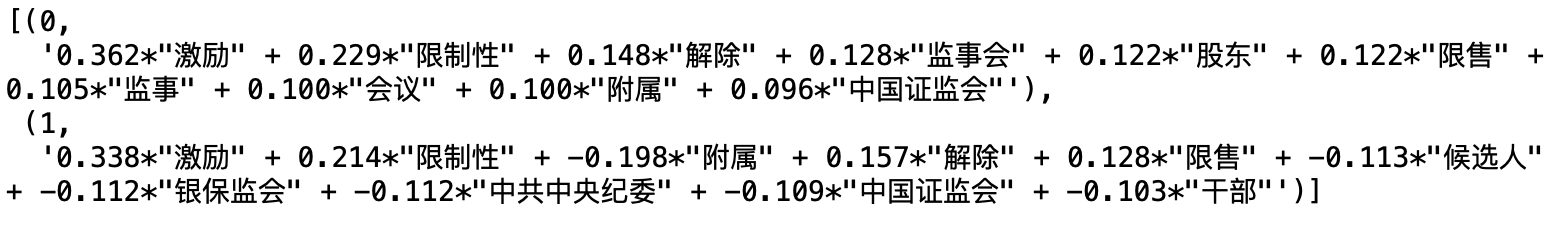


图3:公司公告分布图

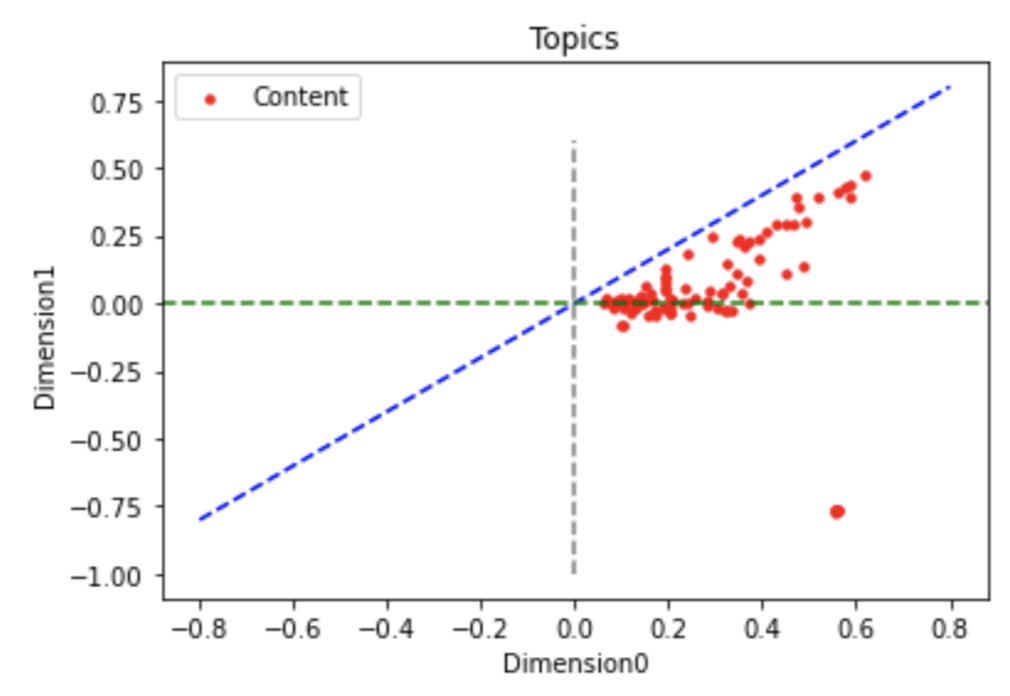
通过对公司公告按照公示日期进行升序排列排列，并绘制如上公司公告分布直方图，从其中不难看出，公司大规模发布公告的主要有三个时间点，一个是4.23，当时正处公司独立董事述职报告发布的时间，多位独立董事同时发表述职报告，同时就公司问题开展相关说明会和监事会会议。一个是5.7，当时公司发布的A股限制性股票激励计划获得国务院国资委批复，并得以面向298位公司高层发放，因而产生颁布十数条对应的股东大会公告、法律文书说明以及独立董事声明等相关公告，该激励计划一定程度上反映了集团在对人才激励上的重视，以及盈利能力的体现。

**Insight3:文档主题分类散点图**

在这里，我利用词袋模型建立话题字典，然后通过LDA模型对已经完成分词的文本数据进行主题分类，通过利用tf-idf对文档进行向量化，为了后续方便利用二维图标进行展示，这里选择两个latent topics的LSI模型，并利用csv中的文档数据进行训练。最终能够将所有文档都化为二维向量，两个向量维度如下所示。



可以看出，两个话题的分类效果一般，两个维度之间的相关性和关键词还是重合度比较高，但是也能够看出一些差异，比如在维度1中比较特别的几个关键词：银保监会、中共中央纪委和干部，说明跟维度0相比，该维度可能更加关注一些企业管理自查以及公权力机构对企业监察的内容，而这部分内容很可能揭示出一些市场应该关注的警示信号。为此，我们将不同的文档依据两个维度的向量绘制如下散点图：



可以看到，不同的文档分别在两个维度上取得了不同分数，通过三条不同的基准线我们能够看到，太钢不锈公司公告在0维度上的得分普遍大于1维度的得分，说明整体而言企业公告揭示的正面信息偏多，公司整体的经营状况良好。但同时也应该看到，存在少部分的企业公告位于右上角，说明其1维度的得分较高，可能说明一定的经营风险或者人事异动，通过公告自查，也发现该部分公告往往意味着股东减持和营收的下降或者管理层人员的非自然调动，因此投资者应该更加关注在右上方的数据点对应的企业公告，。