

不再考虑事件获取数据，根据上市公司的名称

按照结果对新闻进行分类，利好利空中性

文本情感分析

- 词典+规则
 - 词典需要不断丰富
 - 规则->权重
 - 提取文本，分词
- 机器学习

先求有，再求好。一口吃不成胖子。

直接根据文章的标题来确定新闻的分类，不需要打开链接（利好利空中性模糊）粗略的分类

先给出词典和规则

基于词典和规则，机器给出2000条的结果

人工按照自己的词典和规则给出200条的结果

实施步骤

如果目标是完成科技立项结题的话，阶段一和阶段二的工作做完，就ok了。

但科技立项并不是目标，只是把大家聚在一起做事的一个机会。

后期还会顺带参加校长杯等比赛，甚至参加业内的金融科技比赛（有录用offer和实习机会）、深高金的金融科技大赛（获奖可以直接硕士录取，深高金的就业质量高于武中厦）。

我们的最终目标是做出受市场认可的、有应用价值的商业产品。

如果我们做的东西不受市场欢迎，没有落地、没有帮助社会解决一些需求，那这个东西可能毫无意义，除了水创新学分。

阶段三 因子打分的优化

基于规则和词典给出的是离散的分。

如何给出连续的分？

上机器学习。

标注语料哪里来？不断优化、丰富后的基于规则和词典的算法给出的分，可以用来作为机器学习的语料。

阶段2是验证词典和规则的准确性

阶段3是先根据词典和规则生成数据集，然后用数据集来训练模型

阶段四 实体识别

基于规则。

上市公司的数量有限、名称固定，可以使用基于规则的算法。

关键：数据的收集

市场决定是否继续开发

如果市场反应积极，老师会投入更多的资源

人的积极性也会更高