



2022 年“点宽杯”全国高校金融科技黑客松大赛

A 题 基于舆情因子的因子选股策略构建

赛题背景介绍：

据统计，中国股票市场中散户投资者数量约占投资者总数的 95%，中国股票市场也被称为“散户市场”。由于散户投资者的情绪极易受到市场形势与氛围的影响，因此中国股票市场具有很明显的情绪化现象。

伴随着科技的发展，互联网搜索引擎以及各类社交网站不断进入人们的生活，越来越多投资者倾向于在互联网平台交流投资建议与心得，股吧为投资者提供了良好的交流平台，每个人既是信息的制造者，也是信息的传递者。投资者往往在互联网平台上发表个人观点，这一行为会对其他交易者的交易行为产生影响，从而造成股价波动。从投资者情绪的角度出发，构造舆情因子并进行选股策略的构建能够启发投资新思路，为因子选股策略的构建提供更多的选择。

赛题数据描述：

舆情对股价的影响可以从两个方面刻画：股吧评论数量和股吧评论体现的情绪，也称为情感值（分为正向情感和负向情感）。股吧评论数量对股价的影响可能更偏短期；而情感值分析由于涉及了对公司业绩、经济环境的中长期展望等情感，所以对股价会产生持续的影响。本题主要通过分析情感值来构造舆情因子。

深圳点宽网络科技有限公司

舆情因子的构造大致上可以分为两个步骤。首先，我们需要构建一个情感分析模型，这类模型的构建需要使用特定数据集（数据集需包含评论内容和评论情感值）进行训练完成。训练数据集可以从开源网站获取，比如 GitHub 等，而赛题也会提供部分与金融相关的训练数据集。参赛者可自行选择其中的一个或多个，也可以寻找额外数据集进行模型的训练。情感分析模型构建完毕后，参赛者需要自行通过爬虫技术爬取股吧的股票评论，并利用训练好的情感分析模型为每一条评论进行情感赋值。

1. 选定一个或多个股吧，获取股吧的文本数据（例如股票评论），并将获取的数据储存至文件中，格式不限。

要求：自行选择 30 只股票，获取 2021.01 - 2022.06 的数据；

2. 构建情感分析模型，并基于第一问获取的文本数据构造舆情因子；
3. 请验证舆情因子在 A 股市场的有效性；
4. 构建基于舆情因子的单因子选股策略，将第一问挑选出的 30 只股票作为股票池，利用 Auto-trader 平台，在 2021.01 - 2022.06 的时间区间进行回测，并对构建的策略进行分析；

附录 1：赛题数据

训练数据集

因子有效性研报

附录 2：点宽量化平台 AT-edu 软件提交策略说明

点宽杯金融科技黑客松大赛组委会

2022 年 11 月