1. 投资者情绪指标的构建
   1. 数据处理总流程

本文通过爬取网络股评文本、运用最新的深度学习技术对股评文本进行分类、识别股评看涨看跌情绪等步骤，并最终以适当的数学模型构建出股评投资者情绪指标。希望通过这一指标能够较为全面的、实时的反映大部分股市投资者的看涨和看跌的情绪程度。

其中网络股评文本数据全部爬取自东方财富网股吧。据Alexa[[1]](#footnote-1)的最新数据显示，东方财富网全球综合排名位于1707位，中文网站15位，金融财经类网站第1位。而其股吧是国内公认的影响力最大、讨论度最高的股市讨论社区，覆盖了国内较多的投资者人群，其中所蕴含的投资者情绪应能大体反映国内总体的投资者情绪。因此，本文选择东方财富网股吧中的帖子及回复文本作为识别情绪的源文本。

本文构建情绪指标的步骤共可分为三大步，分别是网络股评文本的爬取与预处理、文本情绪的分类与识别和情绪指标的构建。其总体流程如图1所示。

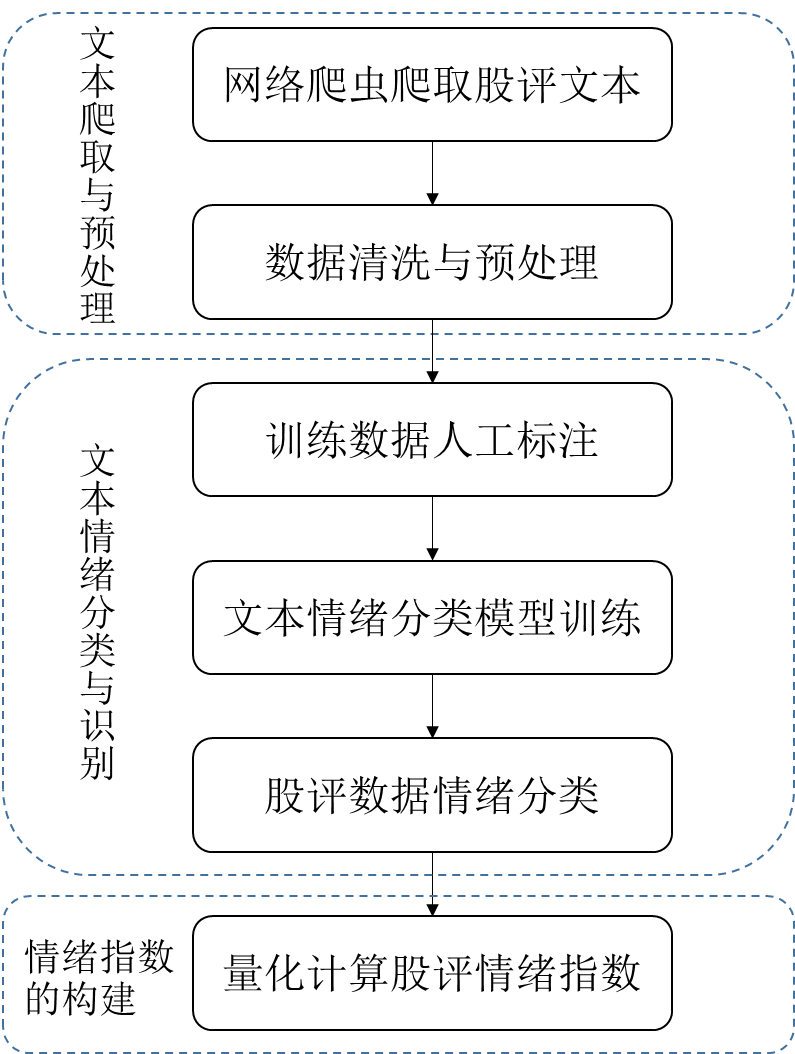


图1 数据处理总流程图

* 1. 网络股评文本的抓取与预处理
  2. 文本情绪的分类与识别
  3. 情绪指标的构建
  4. 数据变量说明

表1 变量名及含义或计算公式说明

|  |  |
| --- | --- |
| 变量名 | 含义或计算公式 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

1. 股评情绪与市场收益率关系的实证研究
   1. 样本选择与数据处理
   2. 描述性统计
   3. 平稳性检验
   4. 股评情绪对股票收益率影响
      1. 股评情绪对月度收益率的影响
      2. 股评情绪对日度收益率的影响
      3. 情绪一致性指数对股票收益率波动的影响
2. 股评情绪对市场交易量影响的实证研究
   1. 描述性统计
   2. 平稳性检验
   3. 股票情绪对市场交易量的影响
   4. 情绪一致性指数对市场交易量的影响

开盘、日内、日间收盘

盘前、盘中、前日

T,t-1,t-2

情绪一致性

1. 股评情绪对股市影响的非对称性效应研究
   1. 看涨和看跌情绪的非对称效应
   2. 极端情绪的非对称效应
2. 股评情绪在资产定价中的横截面效应
   1. 引入股评情绪的四因子模型的横截面检验
   2. 股评情绪对不同股票收益率的影响
   3. 描述性统计分析

对根据以上步骤建立完成的股评情绪指数进行整体性描述，以观察其特征。具体的描述性统计结果如表1所示。

表1 股评情绪指数的描述性统计结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量 | 样本数 | 均值 | 标准差 | 最小值 | 最大值 | 偏度 | 峰度 | Jarque-Bera统计量 |
| 盘前情绪 | 915 | 0.3382 | 0.4673 | -1.3127 | 1.6476 | -0.3473 | 2.8822 | 18.9243 |
| 盘中情绪 | 915 | -0.5793 | 0.3541 | -1.8898 | 0.6974 | -0.1712 | 3.2053 | 6.0780 |
| 前日情绪 | 915 | -0.1963 | 0.4613 | -1.8447 | 1.3782 | -0.0724 | 3.1213 | 1.359239 |
| 盘前情绪一致性 | 915 | 0.0902 | 0.0348 | 0.01953 | 0.2310 | 0.6462 | 3.4676 | 72.0156 |
| 盘中情绪一致性 | 915 | 0.0847 | 0.0302 | 0.0245 | 0.1871 | 0.2950 | 3.0068 | 13.2726 |
| 前日情绪一致性 | 915 | 0.0518 | 0.0186 | 0.0179 | 0.1866 | 1.1114 | 6.7172 | 715.1617 |

表1的统计结果显示，除前日情绪指数之外，所有变量的JB统计量都显著的拒绝正态性分布的零假设。说明各情绪样本的分布不均匀性。其中，盘前情绪样本均值大于0，说明其样本大部分大于0，结合其指标构建公式的含义，可以说明股市休盘后到股市开盘前，网络股评大多有看涨的倾向。而盘中情绪和前日情绪样本则大多低于0，说明盘中情绪和今日开盘前到昨日开盘的整体情绪大多有看跌的倾向，这一结果基本与部慧等[[2]](#endnote-1)一致。此外，三种情绪指数都存在不同程度的左偏，这说明网络股评的相对正面情绪较为集中且一般偏离均值较少，但相对负面情绪较为分散且偏离均值较多。

另一方面，从情绪一致性指数的均值可以看出，盘前的情绪较盘中情绪更为一致。而从情绪一致性指数的最大值可以看出，网络股评的情绪分歧一般比较大，较少出现比较一致的情况。

1. http://alexa.chinaz.com [↑](#footnote-ref-1)
2. 部慧, 解峥, 李佳鸿, & 吴俊杰. (2018). 基于股评的投资者情绪对股票市场的影响. *管理科学学报,* *v.21；No.166*(04), 91-106. [↑](#endnote-ref-1)