4.金融市场因素

4.1股票价格预测

股市预测作为一直以来金融领域研究的热点，搜索、社交媒体、专题论坛等大数据都在股价预测上有所应用，同时也有许多经济学、统计学以及人工智能的相关算法被应用于其中。

4.1.1 股价预测的不同数据来源

在数据来源上，搜索大数据在股票价格预测中得到了一定的使用。[18]用个股公开发行前谷歌搜索量衡量的投资者关注度对于个股市场热销程度、首日超额收益和长期表现有一定的解释力 。也有研究用与谷歌趋势类似的百度指数构建投资者关注度,发现其对股票市场有正向价格影响[19]。

微博等社交媒体舆情对预测股价也有显著作用。近些年,以推特 和微博为代表的网络社交平台上产生用户生成数据具 有空前的规模性和群体性,其中的舆情与社会经济活 动也紧密关联。 例如,通过挖掘微博文本信息构建的舆情指数可有效预测美国股票价格指数变化[26] 。 研究还发现包含观点和情感的微博舆情模型比非舆情微 博模型能更准确的预测股票价格变化,熊市预测比牛 市预测更精确[27] 。 在中国,“牛市冶,“利空冶等股市相 关词汇在微博里出现的频率可以短期预测上证指数变 化,而负面词汇有更强的预测能力[28] 。Bollen, Mao, & Zeng11 发现基于 Twitter(推特)平台表达的公共情绪可以用来预测股市变动。这一研究的数据来源 于两方面：一是 Yahoo·金融发布的道琼斯工业平均指数(DJIA)的收盘价；二是 2008 年 3 月到 10 月，270 万推特用户推送的 970 万条消息。这 970 万条消息经 过情绪评估工具——Opinion Finder 和 GPOMS 被赋值。Opinion Finder 根据文本 内容可以评估“积极”与“消极”两种情绪；GPOMS 根据文本内容可以评估 “calm(冷静)”、“alert(警觉)”、“sure(确信)”、“vital(活泼)”、“kind(美好)”、“happy(高 兴)”六种情绪。结果发现，在道琼斯工业平均指数（DJIA）和 GPOMS 中的“calm （冷静）”情绪之间存在相关性。进一步研究发现， “calm（冷静）”情绪可以很 好地预测道琼斯工业平均指数在未来 2 到 6 天的涨跌情况，而且这种每日预测的 准确率高达到 87.6%。

股票专题论坛则是更加集中了投资者感兴趣的话题和讨论，比传统的社交媒体针对性更强。论坛里有关股票收益的讨论比专业分析师的预测更能体现收益预期,根据这种非正式信息进行操盘的收益也更高[22] 。同时, 股票论坛里异常的发帖活动和观点变化与股票异常成 交量和回报率有显著关系[23] 。用投资者网络讨论信 息构建投资者关注度、乐观情绪与意见分歧指标的研究也显示投资者情绪与 IPO 抑价之间有显著关系[24] 。 从股票论坛中提取中小投资者观点,构建的投资者情绪指标与传统指标存在一定关联性[25] 。

4.1.1 股价预测的不同预测模型

在使用的方法上，最广泛应用的模型之一是线性平稳时间序列自回归模型（AR），例如[1]应用 AR 模型分析中国股指动态收益，[2]将 分解股票收益并利用 AR 模型进行预测。

但股票价格高度非线性和非平稳限制了 AR 模型的适用性，因此机器学习的一些方法得到了研究者的关注。[3]对印度基准指数使用了基于 SVM-KNN 的股票市场趋 势变化分析方法，[4]使用隐马尔科夫模型（HMM）进行股票走势的非线性预测。 随着神经网络逐步发展，在股市预测领域也出现了一些应用，比如[5]对比分析 ARIMA 模型和神经网络在股票价格预测走势，贝叶斯正则化的人工神经网络[6] 也应用于股市预测，[7]结合 SVM、人工神经网络和随机森林提出了一个用于股 价预测的模型。 LSTM[8]能够解决时间序列长短时间依赖的问题，所以在其被提出后得到广 泛应用。RNN 和 LSTM 也在金融预测方面获得了关注，比如[9]结合 RNN 与 AR 预测股市收益，[10]LSTM 通过处理数字和文本数据预测股票价格，[11]LSTM 模 拟交易策略。考虑到股市中存在长期和短期交易者，[12]受到离散傅里叶变换的 启发，提出多频率交易模式 SFM 神经元结构，预测股票长期和短期价格。

除此之外，[13]提取新闻词向量并使用 CNN 对 S&P 500 指数及个股价格进行预测， 另外，他们还使用 NLP 技术处理事件文本信息和知识图提供的实体、属性等信息预测股市波动率。[15]使用混合神经网络以及遗传算法等组成的启发式算法预 测股票价格，[16]使用回归树、自组织映射（SOM）及聚类方法，处理股票收盘价格，利用推荐系统为用户买卖操作提出建议。

4.2 投资市场情绪监测

2012 年，国泰君安推出了“个人投资者投资景气 指数”（简称 3I 指数），该指数系国泰君安研究所对海量个人投资者样本进行持 续性跟踪监测，对账本投资收益率、持仓率、资金流动情况等一系列指标进行统 计、加权汇总后得到的综合性投资景气指数。旨在通过对中小投资人真实投资交 易行为的量化解读，更好地了解投资人对市场的预期以及当前的风险偏好等信息。 此外，基于上市公司经营报表统计、券商投行研报看涨看跌指数（可按地域/行 业/经营领域细分）的分析，也可以为宏观经济运行提供重要参考依据。

4.3 大宗商品期货相关因素

4.4 其它金融相关因素

谷歌趋势中金融相关关键词搜索量可代表投资者搜索股市信息行为密集度。 研究发现金融相关关键词的谷歌搜索量在道琼斯指数涨之前下跌的规律,据此规律建立炒股策略比传统炒股策略收益高[14] 。

岳艳涛, 章雅婷, 张宇,等. 基于大数据分析的橡胶期货交易策略研究[C]// 2015.

商品期货是买卖双方在未来某个约定的日期按约定价格进行交易的标准化协议。期货交易所每秒钟提供两笔交易品种的实时数据,如何对海量数据进行分析并对期货价格做出预测,以期获取稳定的收益成为商业与学术界关注的重点研究问题。本文(1)通过统计聚类等挖掘方法对期货数据进行分析,并对挖掘的指标合理性进行检验;(2)提出周期内的持仓时长概率分布,构造时间与收益的权变函数;(3)利用多种方法对价格波动周期分析,分别得出期货交易价格的短中长三种周期。(4)充分挖掘数据信息,建立综合价格预测与交易模型,并通过实验验证了模型具有很好的效果。首先,根据现有交易数据(成交量、持仓量、总量、买一价、买一量、卖一价、卖一量)确定影响价格的因素。通过成交价与7个因素的散点图和相关系数分析,发现成交价与B1价S1价相关性较高,然后利用ADF单位根和Johansen检验法对数据原序列、一阶差分序列进行平稳性和协整检验,采用Granger因果关系检验确定具有滞后性的影响因素。在对价格的波动方式分类时,首先对折线图分析,利用低通滤波器消除高频随机波动,得到成交价和持仓量波动短周期分别为9.5、11.62min,而成交价中周期为4.51h。最后,通过周期图法与傅里叶级数法,计算2S?在??2最大,即长周期为2天。根据成交价、成交量、持仓量将波动方式分为8类,并对市场交易行情分析。同时创新性提出周期内的持仓时长概率分布,构造时间与收益的权变函数。本文将期货价格的预测分为两部分。首先对橡胶价格进行短期预测,建立自回归与分布滞后模型,用Eviews6.0和SAS9.1软件进行求解和相关性检验。同时利用小波神经网络,对历史数据进行挖掘和模拟分析,进行200和500次训练,预测结果的相关系数达到0.9493,比较预测值与真实值可知模型对短期预测精度高。然后对波动价格长期预测建立阻尼衰减趋势指数平滑模型,并对价格进行预测与误差分析。最后,建立投资收益模型,以期获得最优收益。从期货交易买卖角度对问题进行分析建立模型,利用0‐1规划作为约束,以收益maxTR作为优化目标,通过价格波动预测的结果和实时数据,建立动态规划模型,利用3进制编码(代表每个交易点未交易、买、卖)遗传算法,求解预测价格下的30个交易点的交易策略,以确定每个交易点开空单还是平仓。结论表明,基于期货交易数据的短期分析预测模型与交易策略具有较好的精度和稳定性。

李新（2015）构建反映投资者对国际原油市场关注度的衡量指标，探讨投资者关注与国际原油市场上不同类型的投资者交易持仓关系。采用网络挖掘-特征选择-广义动态因子模型(WM-FS-FHLR)提取互联网搜索关键词，进行特征选择筛选变量，并构建反映互联网关注的“一致指数”。引入生物信息领域的因果分析方法Partial Directed Coherence(PDC)分析因果关系对所构建的指数进行综合评价。该指数是一个基于互联网数据所构建的综合指数，可用于更全面地衡量原油市场中部分投资者的关注，更及时且准确的表征原油市场的动态变化。

进一步构建经济计量模型分析投资者关注对国际原油价格及价格波动的影响程度。构建多组VAR模型来研究投资者关注指数对非商业投资者、商业投资者及散户的持仓关系，从而分析投资者关注指数与投机活动的领先性。