

Design of AI intelligent stock selection system driven by statistical analysis of unstructured data—Take the data of the trade war between China and the United States in April 2025 as an example

Yingying Li

King's College London, London, WC2R 2LS, England

Abstract

As artificial intelligence and big data analysis technologies continue to advance, the application of intelligent stock selection systems in the financial sector is gaining increasing attention. Particularly, through the statistical analysis of unstructured data, AI can assist investors in making quick and accurate decisions in the complex financial market. This paper uses the Sino-US tariff war data from April 2025 as a case study to propose an AI-based intelligent stock selection system design that leverages statistical analysis of unstructured data. The system automatically extracts effective investment signals by deeply analyzing and processing unstructured data in the financial market, integrating machine learning and natural language processing techniques. It achieves intelligent stock selection by constructing predictive models. By exploring the design process, core technology, and application effects of this system, this paper highlights the potential value and application prospects of an AI-driven intelligent stock selection system in modern financial markets.

Keywords

unstructured data, artificial intelligence, intelligent stock selection, natural language processing, 2025 Sino-US trade war

非结构化数据统计分析驱动的 AI 智能选股系统设计 – 以 2025 年 4 月中美贸易战数据为例

李颖颖

King's College London, London, WC2R 2LS, England

摘要

随着人工智能技术和大数据分析技术的不断发展，智能选股系统在金融领域的应用越来越受到关注。特别是通过非结构化数据的统计分析，人工智能可以帮助投资者在复杂的金融市场中快速准确地做出决策。本文以2025年4月中美关税战数据为例，提出了一种基于非结构化数据统计分析的AI智能选股系统设计方案，系统通过对金融市场中的非结构化数据进行深度分析与处理，结合机器学习与自然语言处理技术，自动从中提取出有效的投资信号，并通过构建预测模型实现智能选股。通过对该系统的设计流程、核心技术及应用效果的探讨，本文展示了非结构化数据驱动的智能选股系统在现代金融市场中的潜在价值和前景。

关键词

非结构化数据、人工智能、智能选股、自然语言处理、2025中美贸易战

1 引言

在金融市场中，股市的投资决策一直是投资者关注的核心问题之一。传统的选股方法多依赖于结构化数据，如股票历史价格、交易量、公司财报等。这些数据虽然能够为投资者提供一些参考依据，但在复杂多变的市场环境中，单纯依靠结构化数据往往无法全面反映市场的真实动态。而随着

互联网技术的发展，非结构化数据，如新闻报道、社交媒体评论、财报分析、管理层言论等，成为投资决策中不可忽视的信息源。尤其是通过对这些非结构化数据的统计分析，可以为投资者提供更加细致的市场情绪分析、未来趋势预测和潜在风险评估。

为了弥补传统选股方法的不足，许多研究者和金融机构开始将人工智能和大数据分析技术应用于股票市场的预测与决策中。特别是基于自然语言处理（NLP）技术的非结构化数据分析，能够从大量的文本数据中提取出有价值的信

【作者简介】李颖颖（1992-），女，中国湖南耒阳人，硕士，从事国际管理研究。

息，为股市投资决策提供数据支持。本文设计了一个基于非结构化数据统计分析的 AI 智能选股系统，该系统通过对新闻、社交媒体、财报等多源数据的深度学习与分析，结合股票市场的历史数据，利用机器学习模型进行选股决策。通过该系统，能够有效提升选股的准确性，降低投资风险，实现更加智能化的投资决策。

1.1 贸易战政策冲击与市场初期反应

2025 年 4 月 2 日，美国政府宣布对中国商品加征 34% 的“对等关税”，标志着新一轮贸易战的全面升级。这一政策迅速引发连锁反应，中国国务院关税税则委员会于 4 月 4 日宣布对等反制措施，双方关税博弈在短短一周内从 34% 攀升至 125%。这一系列政策变化对全球金融市场产生了深远影响，特别是对依赖国际贸易的海外出口企业股价造成了显著冲击 [1]。

从市场反应来看，政策公布初期，中美股市均出现剧烈波动。中国上证指数在 4 月 7 日单日暴跌 7.34%，创下 2025 年以来的最大单日跌幅。与此同时，美国标普 500 指数也呈现震荡下行趋势，4 月 8 日跌至 4,982.7703 点的阶段性低点。这种同步下跌反映了市场对贸易战升级的普遍担忧，但随后的市场表现却呈现出明显的分化特征。

1.2 非结构化数据 – 贸易战消息对于中美证券市场成交量及市场情绪指标的意义

成交量数据为理解市场情绪变化提供了重要线索。收

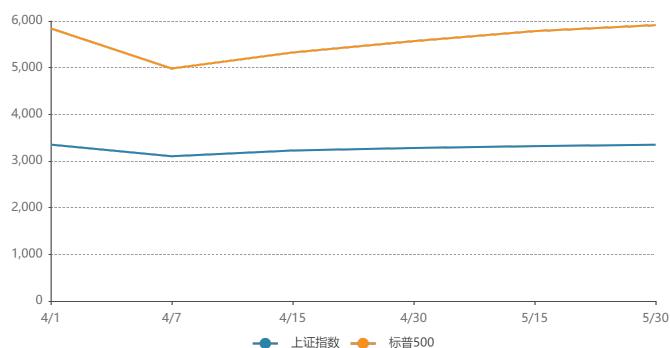
到中美贸易战的影响，中国 A 股市场在 4 月初呈现明显的放量下跌特征，4 月 7 日沪深 300 指数成交量达到 158.92 亿股，较前一交易日增长近 40%，显示恐慌性抛售压力集中释放。进入 5 月后，随着政策对冲措施陆续出台，市场成交量逐渐趋于稳定，但整体仍低于 3 月水平，反映投资者情绪尚未完全恢复。

美国市场则呈现出不同的特征。虽然标普 500 指数在 5 月实现显著反弹，但成交量数据缺失可能暗示市场参与度不足，投资者对贸易战长期影响的担忧仍未消除 [2]。这种量价背离的现象值得警惕，可能预示着市场修复基础的脆弱性。

值得注意的是，5 月 12 日中美宣布暂停部分关税后，市场反应出现分化。阿里巴巴港股在成交量激增至 1.6363 亿股的情况下股价仅短暂反弹，随后继续下行，显示资金逢高减持的谨慎态度。相比之下，波音公司股价在成交量放大后持续攀升，表明市场对美国高端制造业的信心更为稳固。

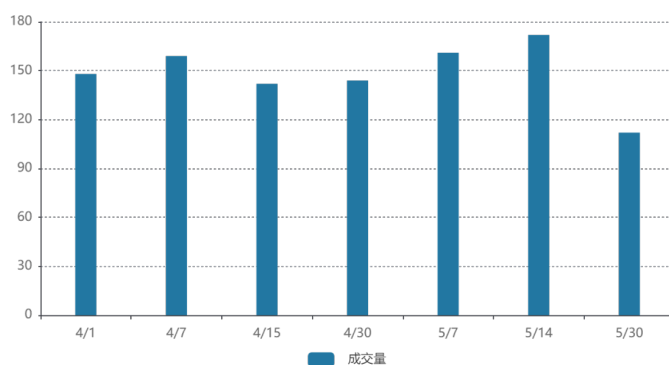
通过分析 4-5 月的股价和成交量数据，可以得出以下主要结论：贸易战对出口依赖型行业冲击显著，但具备全球供应链布局的企业展现出较强韧性；中美市场反应存在明显差异，中国政策对冲效果显著，美国市场修复更快但基础可能较为脆弱；成交量变化揭示了市场情绪演变过程，恐慌性抛售与谨慎反弹并存；贸易战加速了产业结构调整，国产替代和内需扩张成为重要投资主线。

2025年4月主要股指走势对比



(Wind, 2025)

沪深300指数成交量变化 (2025年4-5月)



(Wind, 2025)

2 非结构化数据在股市中的作用与价值

2.1 非结构化数据的定义与特点

非结构化数据是指不遵循传统数据库结构的数据类型，通常包括文本、图像、音频、视频等形式。在股市中，非结构化数据主要指新闻报道、社交媒体评论、公司财报分析、投资者情绪数据等。这些数据通常没有固定的格式，但却包含着大量的有价值信息[3]。与结构化数据不同，非结构化数据往往反映了市场的实时变化和投资者的情绪波动，因此具有较高的分析价值。

非结构化数据的主要特点包括大规模、多样性和不规则性。随着互联网的发展，非结构化数据的生成速度远远超过了结构化数据，其在金融市场中的影响力和应用潜力不断增加。特别是在分析市场情绪、预测股市波动等方面，非结构化数据的价值逐渐被投资者和金融机构所认知[4]。

2.2 非结构化数据在股市分析中的应用 – 以 2025 年 4 月中美贸易战对海外出口企业股价的影响及股市数据分析为例

非结构化数据在股市分析中具有重要的应用价值，能够为投资者提供更为丰富的市场洞察，辅助股市预测和决策。

通过上文标普 500、上证指数及沪深 300 指数对比数据可知，市场情绪分析是非结构化数据在股市分析中的重要应用之一。若在 AI 智能选股系统设计中，加入通过对新闻报道、财经文章、博客和社交媒体评论等文本数据的分析（如 2025 年 3 月已有新闻报道存在中美关税拉锯的风险），能够有效提取出市场的情绪趋势，尤其是与特定股票或行业相关的情绪变化，则可为未来金融市场的稳定性提供有意义的策略因子[5]。情绪分析可以揭示投资者对某只股票的看法和预期，尤其是黑天鹅事件的捕捉，从而为股票价格的涨跌提供一定的预测依据。例如，若市场情绪偏向恐慌，投资者普遍因对未来市场不确定性而带来的恐慌性抛盘，则可能预示着股价将延续下行的趋势。

社交媒体数据也是股市分析中的一个重要来源。投资者在 Twitter、Reddit、股吧等社交平台上的讨论和评论，能够反映出市场对某只股票的关注度及投资者的集体情绪。通过对这些非结构化数据的分析，投资者可以及时捕捉到潜在的投资机会，了解市场热点和趋势，判断股票的潜在投资价值[6]。

此外，结合公司财报数据、行业趋势以及非结构化数据中的市场舆论，分析师可以对股票的基本面进行更为全面的评估。例如，通过分析公司的财报数据、新闻报道以及行业动态，分析师可以判断该公司的经营状况、市场份额以及发展前景，从而提高选股决策的精确度。

总之，非结构化数据的有效分析不仅能够帮助股市投资者更准确地把握市场情绪，还能提供更多维度的信息支持，从而辅助决策，提高投资的成功率。

2.3 挑战与机遇

尽管非结构化数据在股市分析中具有巨大的潜力，但

从海量的非结构化数据中提取有价值的信息依然面临诸多挑战。非结构化数据的来源多种多样，包括新闻文章、社交媒体评论、投资者论坛等，这些数据的格式和内容各不相同，且存在噪声、冗余和不一致性，给数据清洗和分析带来困难。为了提升数据的价值，必须进行有效的预处理，如去除无关信息、填补缺失数据、消除冗余内容等。这一过程对于确保后续分析的准确性至关重要。

此外，传统的分析方法通常难以处理非结构化数据中蕴含的复杂语义和情感信息。例如，在分析股市新闻和社交媒体评论时，情感和语义的准确解读尤为关键，但传统的关键词匹配和简单的统计分析方法无法深入理解文本的情感倾向和语义关系。因此，为了提高分析的准确性，必须依靠先进的自然语言处理（NLP）技术和机器学习模型[7]。自然语言处理技术能够帮助识别文本中的情感倾向、潜在趋势和主题，从而更加精准地提取与股市相关的有价值信息。通过情感分析、主题建模等技术，能够帮助投资者了解市场情绪、预测股价波动，并做出更有依据的投资决策。

总的来说，尽管非结构化数据的处理面临挑战，但通过有效的预处理和先进的分析技术，能够充分挖掘这些数据中的潜在价值，从而为股市分析提供更深刻的洞察和预测能力。

3 AI 智能选股系统的设计与实现

3.1 系统架构设计

本系统的核心思想是结合非结构化数据的统计分析与人工智能算法，构建一个自动化的选股平台。该平台通过多种数据源和先进的分析技术，帮助投资者根据实时市场动态做出精准的选股决策。系统主要由以下几个模块组成：数据采集与预处理模块、情感分析与特征提取模块、机器学习模型训练模块、预测与选股模块以及可视化展示模块。

以 2025 年 4 月 -5 月中国及美国股市的成交量以及以及股价的数据分析为例，美国对华采取的关税趋势，是可通过新闻媒体、社交媒体等非结构化数据统计分析，并采取自然语言处理（NLP）技术和机器学习模型及通过词频抓取达到分析及为投资市场及时发出警示及建议，辅助选股系统有效降低失败率，从而提高投资收益率。

数据采集与预处理模块是整个系统的基础，负责从新闻、社交媒体、公司财报等多渠道获取实时的非结构化数据，并进行数据清洗、去噪和结构化处理。通过处理后，数据可以更好地为后续分析提供准确的信息来源。情感分析与特征提取模块则使用自然语言处理（NLP）技术分析文本数据，提取其中的情感信息和市场趋势。此模块能够识别新闻和社交媒体内容中的情绪倾向，从而了解市场对某只股票的看法，为选股决策提供情感基础。

机器学习模型训练模块则使用历史股市数据进行训练，构建股票市场的预测模型。通过学习历史数据中的趋势和模式，模型能够对未来市场走势做出预测，并对不同股票进行评估。

机器学习模型训练模块设计是非结构化模型其中一个重要的一个环节，若所设计的策略因子中，希望加强对中美间的贸易摩擦的识别，可加入此次中美贸易战的历史数据作为机器学习模块中的一环进行训练，以对未来遇到类似情景，通过 AI 模型对投资决策及板块提供有意义的选股模型。

参考例子如下：

2025 年 4 月至 5 月期间，美国股市整体成交量和股价波动较大。以下是部分代表性企业的数据：

可以看出，部分科技类企业如特斯拉、英伟达、微软等在 4 月至 5 月期间股价上涨，而一些传统行业如医疗健康

(UNH.N) 则出现了较大的跌幅。这反映出市场对高科技、AI、新能源等领域的偏好，而对受贸易摩擦影响较大的行业则持谨慎态度。

2025 年 4 月至 5 月，中国 A 股市场整体表现疲软，但部分内需主导的行业如食品饮料、医药、军工等板块表现相对较好。以下是部分 A 股企业的涨跌幅数据：

从数据来看，中国 A 股市场在 4 月整体呈现下跌趋势，上证指数下跌 1.7002%，深证综指下跌 4.4631%。但部分内需主导的行业如食品饮料、医药等板块表现相对较好，显示出市场对内需的偏好。

数据采集 1. 美国股市成交量与股价变动

Wind 代码	证券简称	成交量 (股)	成交额 (亿元)	涨跌幅 (%)
TSLA.O	特斯拉 (TESLA)	51.5412 亿	14,711.5566	33.6858
NVDA.O	英伟达 (NVIDIA)	115.7025 亿	13,204.4779	24.6817
AAPL.O	苹果 (APPLE)	28.0321 亿	5,587.2053	-9.4614
PLTR.O	PALANTIR TECHNOLOGIES	44.2592 亿	4,694.62	56.1374
MSFT.O	微软 (MICROSOFT)	10.4011 亿	4,247.7023	22.8603
AMZN.O	亚马逊 (AMAZON)	22.5046 亿	4,236.7607	7.7525
META.O	脸书 (META PLATFORMS)	7.4387 亿	4,219.3381	12.3412
GOOGL.O	谷歌 (ALPHABET)-A	17.8546 亿	2,840.6558	11.0579
UNH.N	联合健康集团 (UNITEDHEALTH)	8.1115 亿	2,825.1268	-42.3561
MSTR.O	微策投资	6.8666 亿	2,385.8698	28.0258

(Wind,2025)

数据采集 2. 中国 A 股成交量与股价变动

Wind 代码	证券简称	涨跌幅 (%)
301209.SZ	联合化学	159.0884
601086.SH	国芳集团	117.0996
002165.SZ	红宝丽	104.2693
001212.SZ	中旗新材	85.7192
920106.BJ	林泰新材	84.9448
300972.SZ	万辰集团	82.7915
603696.SH	安记食品	76.6908
600644.SH	乐山电力	71.6867
300652.SZ	雷迪克	70.1998
002915.SZ	华光源海	65.4235

(Wind,2025)

数据采集 3. 中美股市整体表现对比

Wind 代码	证券简称	涨跌幅 (%)
IXIC.GI	纳斯达克指数	0.85
DJL.GI	道琼斯工业指数	-3.1722
000001.SH	上证指数	-1.7002
000300.SH	沪深 300	-3.003
881001.WI	万得全 A	-3.1503
HSI.HI	恒生指数	-4.3261
399106.SZ	深证综指	-4.4631

(Wind,2025)

从整体来看，美国股市在4月表现略优于中国A股，纳斯达克指数微涨0.85%，而上证指数下跌1.7002%。这可能与美国科技企业在全球贸易摩擦中展现出更强的抗压能力和市场预期有关。

如上所述，通过机器学习对此次事件历史数据进行模型训练，结合预测与选股模块基于训练好的模型输出，未来遇到相似情景，通过AI甄别，可迅速响应，自动进行选股决策，帮助投资者从众多股票中筛选出潜力股。最后，以可视化展示模块将分析结果以直观的方式呈现给用户，通过图表、趋势线等形式，帮助投资者理解分析结果，做出更加明智的投资决策。

该系统通过集成自然语言处理、机器学习和数据可视化技术，为投资者提供了一种高效、智能的选股工具，大大提高了投资决策的准确性和实时性[8]。

3.2 非结构化数据的情感分析

上文通过结构化数据采集对比，分析关联性及其有效性，证明对于突发事件、大事件与证券市场的量价行成有效的因子。在现在大数据时代，除了使用机器学习模块训练历史数据的捕捉及识别逻辑外，对于媒体报道的情感识别，也是非结构化数据统计分析驱动中另一重要的环节。

情感分析是自然语言处理(NLP)中的一项关键技术，它通过分析和识别文本中的情感倾向来量化公众和投资者对某个主题、事件或对象的态度。在股市分析中，情感分析可以帮助投资者了解市场的情绪变化，从而做出更为精准的选股决策。特别是在分析新闻报道、社交媒体评论、投资者论坛等非结构化数据时，情感分析能够揭示出市场情绪的动向，为股市预测提供强有力的数据支持。

例如，通过对与某只股票相关的新闻报道和社交媒体评论进行情感分析，系统能够识别文本中的正面、负面或中性情感[9]。如果某只股票的新闻报道中充满了负面情绪，如公司管理问题、财务危机或市场对其前景的质疑，这可能意味着该股票在短期内会受到负面市场情绪的影响，投资者应谨慎对待。反之，如果舆论情绪呈现积极态度，公众对该股票的未來前景持乐观态度，那么该股票可能会吸引更多的投资者关注，并迎来市场的热捧。

这种解释代入例子中，可参考对于2025年4月面对贸易战冲击，中美两国采取了不同的应对策略，并通过各大媒体报道，这对股市表现产生了深远影响。如：中国政府密集出台内需刺激政策，包括城市更新行动、房贷利率下调等措施，有效缓冲了出口下滑的压力。这些政策推动商贸零售、食品饮料等内需主导型行业在4月下旬至5月表现突出，成为稳定市场的重要力量。美国则通过关税豁免和供应链调整来缓解冲击。对部分商品的90天关税暂缓政策，以及东南亚供应链的关税豁免优势，使得部分美国企业能够快速调整全球布局。这种灵活的政策应对在一定程度上解释了为什么美国股市在5月能够实现较快修复。

从行业结构来看，贸易战加速了全球产业链的重构进程。半导体设备国产替代、农产品自给率提升、纺织业产能转移等趋势都在股价表现中得到体现[10]。

通过自然语言处理(NLP)中的情感识别，加入该结构性变化的趋势，不仅能有效提高AI选股策略对于短期市场波动的抓取能力，更能通过此因子加入更多的决策因子如通过对不确定及反复的政客博弈、冷摩擦的报道，从而结合量化提高策略的胜率。

美国出口导向型行业		中国A股行业	
电子元件	如亿鹏能源(ELPW.O)、奥斯汀光电科技(OST.O)等，4月股价涨幅显著，分别为315.7667%和115.7895%。	食品饮料	如国芳集团(601086.SH)、安记食品(603696.SH)等，股价上涨明显。
电池与储能	如Microvast(MVST.O)、SES AI(SES.N)等，股价也有明显上涨，分别为64.9573%和72.6138%。	医药	如科兴制药(688136.SH)、一品红(300345.SZ)等，股价也表现强劲。
工业机械	如芝加哥机械(CVR.A)、高频电子(FEIM.O)等，股价也有所上升，分别为21.6713%和18.8498%。	军工	如林泰新材(920106.BJ)、雷迪克(300652.SZ)等，受益于政策支持，股价上涨。

(Wind, 2025)

从此数据看来，此次影响对于美国的电子元件、电磁储能与工业机械以及A食品饮料、医药和军工带来消息面的股价红利。

在具体实现过程中，情感分析通常采用深度学习模型来提高分析的准确性。常见的模型包括LSTM(长短期记忆网络)和BERT(双向编码器表示转换器)。LSTM能够有效处理和记忆文本中的长期依赖关系，适用于捕捉情感在长文本中的变化，而BERT则是基于Transformer架构的预训练模型，能够更好地理解上下文中的语义信息。通过对这些模型的训练，系统可以自动识别文本中的情感倾向，并根据情感的强度对数据进行量化，从而为投资者提供更加精准的

舆情分析和选股决策依据。

3.3 机器学习模型与选股决策

基于非结构化数据提取的特征，机器学习模型可以用于预测股票的未來表现。在训练过程中，系统将历史的股票数据与新闻情感、社交媒体情绪、财报数据等特征进行结合，训练出一个能够预测股票涨跌幅度的模型。常见的机器学习算法如随机森林、支持向量机(SVM)、深度神经网络(DNN)等都可以应用于此任务[11]。

经过训练后的模型能够根据实时获取的非结构化数据，预测股票未来的走势，并自动给出选股建议。例如，当模型分析到某只股票的情感倾向为正，并且相关财报数据表现良

好时，系统会向用户推荐该股票作为潜力股。此外，模型还可以根据用户的风险偏好和投资目标，提供个性化的投资组合推荐。

4 AI 智能选股系统的优势与挑战

4.1 系统的优势

AI 智能选股系统相比传统的选股方法，具有显著的优势。首先，通过对大量非结构化数据的分析，系统能够实时捕捉到市场动态和投资者情绪的变化，为选股提供更加全面的依据。其次，AI 系统通过机器学习不断优化和自我调整，可以根据历史数据和市场变化进行精准预测，减少人为决策带来的偏差。此外，系统能够处理大量的数据，并且能够自动执行选股任务，提高了选股效率，减少了人工干预。

4.2 面临的挑战

尽管 AI 智能选股系统具有诸多优势，但在实际应用中仍然面临一些挑战。首先，非结构化数据的质量和来源多样，如何有效筛选和清洗数据，去除噪声，确保数据的准确性是一个难点。其次，情感分析和特征提取的准确性直接影响到系统的预测能力，因此如何提高自然语言处理模型的精度和鲁棒性也是一个关键问题 [12]。最后，股市的变化具有较大的不确定性，AI 选股系统虽然可以提供一定的参考依据，但仍然无法完全消除市场的风险。

5 结语

AI 智能选股系统的设计与应用，充分展示了非结构化数据分析在现代金融市场中的巨大潜力。通过引入先进的自

然语言处理技术和机器学习算法，系统能够从复杂的市场数据中提取有效信息，进行精确的股市预测，帮助投资者做出科学的投资决策。尽管在实际应用中仍然面临数据质量、算法准确性等挑战，但随着技术的不断进步，AI 选股系统将不断优化，提升选股的准确性和效率，为投资者提供更强大的数据支持和决策依据。

参考文献

- [1] 龚振庭,陈妍蓓.碳交易市场、能源市场与低碳股票市场的风险溢出效应——基于联合溢出指数模型的实证研究[J].新疆财经,2023,(05):16-28.
- [2] 秦彦焱.碳市场、股票市场与原油期货市场间的波动溢出效应研究[D].西安理工大学,2023.
- [3] 许超.动量视角下股票市场价格变动趋势研究[D].北京交通大学,2023.
- [4] 蔡川奇.时频域视角下的我国碳市场与股票市场的溢出效应研究[D].云南财经大学,2023.
- [5] 齐洪杰.疫情冲击视角下我国碳市场与股票市场的相依性研究[D].南京信息工程大学,2023.
- [6] 2025年4月到5月美股成交量、成交额及涨跌幅 (Wind,2025)
- [7] 2025年4月中美贸易战升级后美国对中国出口企业股价涨跌幅 (Wind,2025)
- [8] 2025年4月中美股市涨跌幅 (Wind,2025)
- [9] 2025年4月中国A股涨跌幅 (Wind,2025)
- [10] 2025年4月投资展望:好事多磨,未来可期 (雪球, 2025)
- [11] 出口对A股的影响与分析 (搜狐, 2025)
- [12] 财务BP在企业风险管理中的作用与面临的挑战 刘勇; 133_135