

基于 Transformer 和大语言模型生成的公式化阿尔法因子的情感感知股票价格预测

日本神戸兵库县立大学信息科学研究生院

af24a008@gsis.u-hvogo.ac.jp

日本神戸市兵库县立大学信息科学研究生院

kawashima@gsis.u-hvogo.ac.jp

摘要

传统上，交易员和量化分析师通过领域专业知识和反复试验，手动构建公式化阿尔法（即识别金融数据中模式或信号的数学表达式）来应对阿尔法衰减问题。这一过程往往耗时且难以规模化。随着大型语言模型（LLMs）的最新进展，如今借助 LLMs 的推理能力，实现此类阿尔法的自动化生成已成为可能。本文介绍了一种新颖的框架，该框架将基于提示词的 LLM 与 Transformer 模型相结合，用于股票价格预测。LLM 首先利用结构化输入（如历史股票特征（收盘价、开盘价、最高价、最低价、成交量）、技术指标、目标公司及相关公司的 sentiment 得分）生成多样化且具有适应性的阿尔法。这些阿尔法并非直接用于交易，而是被视为捕捉金融数据中复杂依赖关系的高级特征。为评估这些由 LLM 生成的公式化阿尔法的有效性，随后将这些阿尔法特征输入到 Transformer、LSTM、TCN、SVR 和随机森林等预测模型中，以预测未来股票价格。实验结果表明，LLM 生成的阿尔法显著提高了预测准确

关键词 大语言模型 · 股票价格预测 · 情感分析 · Transformer · 深度学习

1 引言

金融市场是复杂的系统，其价格受公司基本面、技术形态、市场情绪和外部事件的影响。在金融领域和机器学习领域，预测股票价格仍然是一项重大挑战。回归和基于树的方法等传统模型往往难以捕捉金融时间序列

深度学习模型，尤其是 Transformer 模型，最近在序列建模任务中展现出了优异的效果。它们学习长程依赖关系的能力使其适用于价格变动预测 [1, 2, 3]。然而，仅使用原始金融特征往往会导致预测结果存在噪声且不稳定。为了提升模型性能，许多研究人员和从业者会创建名为“阿尔法因子”或“阿尔法”的特

$$\text{Alpha}_{\{t\}} = \left(\frac{P_{\{t\}} - \text{SMA}_{\{20, t\}}}{\text{SMA}_{\{20, t\}}} \right) \cdot \log \left(1 + \text{Sentiment}_{\{t\}} \right)$$

该阿尔法因子结合了技术指标和情绪信息。术语 $P_{\{t\}}$ 表示

当前价格 P_t 与其 20 日简单移动平均线的百分比偏差，用于指示短期超买或超卖状况。对数情绪项 $\log(1 + \text{Sentiment}_t)$ 会根据当日市场情绪调整这一信号，其中较高的情绪值会放大该信号，而负面或中性情绪则会减弱该信号。通过结合技术指标和情绪数据，这一阿尔法因子旨在比单独使用技术指标或情绪数据更有效地捕捉市场信号。在实际操作中，阿尔法因子的有效性并非永久存在。随着时间的推移，当越来越多的交易者发现并利用某一阿尔法因子时，它往往会失去预测能力。这种现象被称为阿尔法衰减 [4]。阿尔法衰减是量化金融领域

传统上，新阿尔法因子的设计一直是一个手动过程，由人类的直觉和经验驱动。分析师和量化分析师研究历史模式、测试假设，并通过反复试验来完善公式。然而，这种方法速度慢、成本高，且受限于人类的创造力。近年来，像 GPT 这样的大型语言模型（LLMs）在推理、总结和创意生成方面展现出了强大的能力。在这种情况下，基于提示词的大语言模型（LLMs）具有多个优势。首先，它们能够按照清晰的指令（提示词）快速生成大量候选公式。其次，生成的阿尔法因子通常具有可解释性，因为它们是以人类可理解的形式编写的，这与深度神经网络所学习的隐藏特征不同。第三，基于提示词的大语言模型具有灵活性。通过调整提示词设计，用户可以引导大语言模型关注市场的不同方面，例如动量、均值回归或情绪。这种灵活性使得

在本文中，我们提出了一个新框架，该框架利用基于提示词的大语言模型（LLMs）从股票特征、技术指标和情绪得分中自动生成公式化阿尔法因子。这些生成的阿尔法因子作为额外输入提供给基础 Transformer 模

继我们之前的研究 [6]，我们还将相关公司的情感作为输入特征之一。相关公司指的是与目标公司存在关联的公司，例如业务合作伙伴、供应商或竞争对手。原因是这些公司的新闻情感也会影响目标公司的股票走势。在我们早期的研究中，我们应用命名实体识别（NER）从有关目标公司的新闻文章中提取公司名称。其基本假设是，与目标公司频繁在同中现左同一管立音由的公司可能存在某种形式的业务关系或关联性。

总之，本文解决了以下两个研究问题（RQ）：

- 研究问题 1：能否将相关公司的情感信息有效整合到由大语言模型生成的阿尔法公式中？
- 研究问题 2：利用 Transformer 等深度学习技术时，具有情感感知的大语言模型生成的阿尔法信号是否能提

本文其余部分的结构如下。首先，列出相关工作（第 2 节）。然后描述研究方法（第 3 节）。第 4 节呈现实验结果。第 5 节对研究发现进行进一步讨论。最后，第 6 节总结全文。

2 相关工作

2.1 基于机器学习模型的股票价格预测

预测股票价格吸引了大量的研究关注。传统方法包括 ARIMA [7] 和 GARCH [8] 等统计模型。这些模型假设数据中存在特定模式，例如平稳性或恒定波动性，但实际的金融数据往往违背这些假设。因此，这些模型在捕捉市场动态方面存在局限性。近年来，机器学习模型，特别是深度学习模型，在股票价格预测中取得了显著进展。这些模型能够处理更复杂的数据模式，并捕捉到传统统计模型难以识别的非线性关系。然而，机器学习模型通常面临过拟合和解释性差的问题，这使得它们在金融领域的实际应用仍然具有挑战性。

随着机器学习的兴起，线性回归 [9]、支持向量机 [10]、随机森林 [11] 和梯度提升 [12] 等模型已被广泛用于股票预测。与传统的统计方法相比，这些模型能更好地捕捉非线性模式。然而，它们仍然需要精心的

深度学习方法在股票预测方面展现出了巨大潜力。循环神经网络（RNNs）和长短期记忆（LSTM）网络是最早用于金融时间序列的深度学习模型之一 [13, 14, 15]。它们旨在捕捉数据中的时间依赖性。虽然 LSTM 通过

后来出现的 Transformer 通过依赖自注意力机制克服了这些限制。Transformer 能更有效地捕捉长期依赖关系，已成为时间序列预测任务的热门选择，在金融领域也不例外。多项研究通过将历史价格、技术指标和其他特征作为序列输入，对 Transformer 架构进行了调整，以用于股票预测 [16, 17]。一些方法对注意力机制进行了修改，以便更好地聚焦于重要的市场事件 [18]。还有一些方法将 Transformer 与卷积神经网络

2.2 特征工程与公式化阿尔法因子

特征工程仍然是构建成功的股票预测模型的关键部分。许多研究人员提出将移动平均线、相对强弱指数（RSI）、指数平滑异同平均线（MACD）和布林带等技术指标作为特征 [21, 22, 23]。一些研究还使用基本面数据 [24, 25]，如市盈率、市净率等来构建特征。然而，如何有效地结合这些特征仍然是一个挑战。阿尔法因子的概念是量化金融的核心。早期研究基于动量、均值回归和价值投资等理论手动构建阿尔法因子 [27]。后来，研究转向通过使用遗传算法和符号回归来搜索有利可图的公式，从而实现阿尔法因子发现的自动化 [28]。最近，一些研究提出使用强化学习来生成或选择阿尔法因子，以减少人为偏差并提高性能 [29, 30]。尽管做出了这些努力，阿尔法因子仍然是一个长期存在的问题。因为曾经强大的阿尔法因子在公开市大型语言模型（LLMs）为金融领域开辟了新的可能性。LLMs 已被用于总结 earnings 电话会议、从新闻文章中预测市场情绪，甚至生成交易策略 [31]。然而，使用 LLMs 生成公式化阿尔法因子仍是一个新领域。基于提示词的方法允许用户通过设计好的指令来控制 LLMs 的输出。这种方法可以引导 LLM 创建结构化且可解释的公式，而非自由形式的文本。它结合了人类的创造力和机器的效率。因此在阿尔法因子生成方面与 [5] 等先前的研究不同，那些研究使用阿尔法生成直接的交易信号，而本研究将阿尔法视为高级结构化特征，这些特征捕捉了金融输入之间的复杂关系。然后，这些结构化特征被用作预测模型的输入，以预测股

3 方法

本节阐述了本研究中使用的数据源、预处理步骤、模型设计和训练方法。我们旨在结合传统金融特征、sentiment 分析以及基于提示词的大语言模型生成的阿尔法因子来预测股票价格。

图 1 展示了本研究提出的框架。首先，大语言模型（LLM）利用股票特征、技术指标等基本特征，以及相关公司的情感得分，生成公式化阿尔法因子，然后将其输入 Transformer 模型以预测次日的股票价格。

3.1 数据

五家公司（苹果、汇丰、百事、丰田和腾讯）的股票数据是使用 yfinance Python 库收集的。这些数据涵盖的时间段为 2016 年 5 月 1 日至 2024 年 5 月 8 日。选择这些公司是为了涵盖不同的行业和市场。技术指标是使用 pandas-ta 库计算的。技术指标是基于股票价格、成交量或持仓量的数学计算。它们被交易者广泛用于识别市场趋势和预测未来价格走势。表 1 显示了本研究中使用的技术指标。简单移动平均线



图 1：所提方法的整体图景

（SMA）通过计算固定天数内的平均收盘价来平滑短期价格波动。它有助于识别一段时间内的基本趋势。指数移动平均线（EMA）对近期价格赋予更大权重，因此相比 SMA，它能更快地对新的市场变化做出反应。动量指标通过将当前价格与几天前的价格进行比较，来显示价格变化的速度，它能预示价格变动的强度。相对强弱指数（RSI）衡量价格变化的速度和幅度，有助于判断股票是否处于超买或超卖状态。移动平均收敛散度（MACD）追踪两条 EMA 之间的差值，它能凸显趋势强度和方向的变化。布林带在移动平均线周围设定上

表 1. 所用技术指标列表

除了股票数据外，还利用财经新闻推送和股票新闻情感数据 API¹ 从雅虎新闻等网站下载了与这五家公司及其关联公司相关的每日新闻文章。根据我们之前的研究，关联公司被定义为在关于这五家目标公司的新闻文章中频繁出现的公司 [6]。例如，比亚迪在丰田的新闻文章中出现了一百多次，这表明两家公司之间存在联系。新闻文章有助于捕捉市场情绪，而市场情绪可能对股票价格产生重要影响。此前，我们使用命名实体识别（NER）从与目标公司相关的新闻文章中提取实体名称，并基于其出现频率进行统计。然后，所有识别出的目标公司按出现频率由高到低排序，对于每个目标公司，选取提及频率最高的前五家公司作为其关联公司。在本研究中，提取出的相关公司将由大语言模型进一步分析，以确定那些与股票走势有密切关联且最具影响力的公司。随后，大语言模型选出的公司将被纳入公式化阿尔法因子中。

表 2. 相关公司 [6]

每篇新闻文章都会经过处理以计算极性得分。极性得分是使用 VADER（valence Aware Dictionary and sEntiment Reasoner，情感感知词典与推理器）[32] 计算得出的，这是一种基于词典和规则的情感分析工具，专门用于分析社交媒体和短文本中表达的情感。它通过使用一个预定义的词典（词汇表）来工作，该词典将词语映射到它们的情感强度得分，包括积极、消极或中性。极性得分用于衡量一段文本的积极、中性或数据的前 70% 用于训练，剩下的 30% 用于测试。对输入特征采用 5 天的滑动窗口。

3.2 利用大语言模型生成公式化阿尔法因子

为了生成新的交易公式，使用了一个名为 deepseek-r1-distill-llama-70b 的大语言模型。该模型由美国公司 Groq² 开发，是一个更大模型的蒸馏版本。知识蒸馏是一种技术，通过这种技术，较小的模型（学生模型）学习模仿更大、更强大的模型（教师模型）的行为。这个过程在保留原始模型大部分性能的同时，减小了模型的规模并降低了计算成本。大语言模型（LLM）通过提示词进行控制。在本研究中，提示词（表 3）引导大语言模型生成公式化阿尔法因子。具体而言，该提示词是一个结构化的指令任务提示词，旨在指导模型生成用于股票价格预测的公式化

¹<https://eodhd.com/financialapis/stockmarket-financialnews-api>
² <https://groq.com/>

表 2. 阿尔法生成的结构化提示词

任务提示：生成 {公司} 股票价格的预测阿尔法因子

目标：生成公式化的阿尔法信号，以预测 {公司} 的股价，所使用的数据包括：1. 股票特征（例如，收盘价、开盘价、最高价、最低价、成交量）；2. 技术指标（例如，相对强弱指数（RSI）、移动平均线、指数平滑异同平均线（MACD））；3. 目标公司及其相关公司（例如，竞争对手、业务合作伙伴、供应商）的情绪数据。

输入数据：一个单独的 pandas.DataFrame（JSON 格式的 DataFrame 输入），其中行代表交易日，列包括：

-

- 股票特征：开盘价、收盘价、最高价、最低价、成交量

而

1. 阿尔法公式：提出 5 个结合股票特征、跨公司情绪差异或技术指标的公式化阿尔法。

~

示例 Alpha: $\alpha_1 = \text{苹果_5D_收益} + 0.5 \times (\text{苹果_情感} - \text{谷歌_情感})$

确保模型理解领域背景。该提示词还包含一个具体示例，以设定预期的输出格式。

大语言模型（LLM）的输入包括技术指标、情绪分数和基本股票特征。输出是新的阿尔法公式，这些公式随

3.3 基于 Transformer 的股票预测

这些由大语言模型生成的阿尔法特征被输入到一个编码器 - 解码器 Transformer 模型中（图 2）。该

具体来说，在编码器端，输入包括五个公式化的阿尔法因子，采用五天的滑动窗口。这些因子通过一维卷积嵌入层处理，以发现随时间变化的有用模式。同时，日期、星期、月份等与时间相关的特征由时间嵌入层处理，帮助模型理解时间周期。这两个层的输出与位置嵌入相结合——位置嵌入有助于模型记住数据的顺序

在解码器端，输入是带有五天滑动窗口的收盘价。它们会经历相同的处理过程：首先通过一维卷积嵌入来学习模式，然后是用于时间特征的时间嵌入，最后添加位置嵌入。解码器会同时利用自身输入和编码器的输出来预测次日的收盘价。这种设置有助于模型同时学习价格走势和时间模式，从而做出更准确的预测。该模型使用预测收盘价与实际收盘价之间的均方误差（MSE）损失函数进行训练。模型采用 Adam 优化器进行优化，学习率为 5×10^{-5} ，训练轮次为 50 轮。批次大小因股票而异，范围在 8 到 64 之间。关键超参数（如学习率和批次大小）是根据验证性能通过经验迭代选择的。研究了一组常用估 是效选择的配置并测试了

3.4 用于性能比较的模型

在本研究中，加入了一些股票预测中常用的模型进行比较。以下是本研究使用的每个模型的简要说明。



图 2. Transformer 结构 [6]

- 长短期记忆网络（LSTM）是一种循环神经网络，擅长学习时间序列数据中的模式。它利用记忆单元和门控机制来保留重要信息，并随着时间的推移丢弃不太有用的信息。
- 时间卷积网络（TCN）是一种深度学习模型，它使用一维卷积层处理序列数据。与循环神经网络不同，它以并行方式处理序列，并通过扩张卷积捕捉长程依赖关系。
- 支持向量回归（SVR）是一种机器学习方法，它能找到一条线或曲线来拟合数据，同时将误差控制在设定的范围内。它在中小型数据集上表现良好，并且可以处理线性和非线性趋势。
- 随机森林是一种集成模型，它构建多个决策树并结合它们的结果。通过对单个树的输出进行平均，它能减少过

3.5 评估

均方误差（MSE）用于评估模型生成的股票价格预测的准确性。它衡量的是预测股票价格与实际价格之间的平均平方差。MSE 值越低，表明预测值越接近真实值，意味着预测准确性越高。MSE 的计算公式如下。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

其中， y_i 是实际股票价格， \hat{y}_i 是预测股票价格， n 是预测的总数量。

此外，为了凸显纳入目标公司及其关联公司情感的创新性，我们评估了将情感信息整合到 Transformer 模型中用于股票价格预测的影响，这将在 4.3 节中进行说明。对于不包含情感的基准场景和包含情感的实验场景，我们对每只股票进行了十次独立运行。选择这个次数是机器学习实证研究中的常见做法，旨在平衡统计稳健性和计算效率。我们报告了每种情况下的平均均方误差（MSE）和标准差，以体现模型性能并解释训练过程中的变异性。这确保了比较的公平性和统计合理性，使我们不仅能够评估情感特征带来的平均性能提

结果

表 4 展示了所有五家公司的由大语言模型生成的公式化阿尔法因子。表 6 展示了 Transformer 在输入包含和不包含情感分数情况下的性能。图 3 展示了大语言模型为丰田公司生成公式化阿尔法因子的推理过程。表 5 展示了使用表 4 中的公式化阿尔法因子对所有公司进行预测时的均方误差（MSE）表现。图 4 可视化了所有模型在汇丰银行股票预测中的预测性能。从该图表可以明显看出，当使用大语言模型生成的特

4.1 阿尔法生成

根据图 3，DeepSeek 模型通过模拟类人推理过程生成公式化阿尔法因子，该过程系统地整合了技术指标和基于情感的特征。推理过程由大语言模型自动生成。具体而言，该模型首先回顾量化交易中阿尔法公式的定义和目的。然后，它策略性地结合各种输入，例如动量指标（如 Momentum_3、Momentum_10）、震荡指标（如 RSI_14）、趋势跟踪指标（如 SMA_5、SMA_20）、基于成交量的指标（如 OBV），以及通过情感分析得出的公司特定极性分数。每个公式都旨在捕捉一种独特的市场动态，例如动量与情感的一致性、被积极情

通过这一推理过程，DeepSeek 模型使分析师和交易员能够理解每个预测信号背后的原理，而不是依赖黑箱输出。这种透明度有助于更好地进行风险评估，因为用户可以评估每个阿尔法所捕捉的逻辑和市场状况。此外，通过系统地整合各种指标，这一过程有助于发现不明显的关系，并助力构建更稳健、更多元化的交易策略。例如，对于百事公司（见表 4），大型语言模型捕捉到的一个不明显关系出现在 α_1 中：

$$\alpha_{1,t} = Momentum_{3,t} + Amazon_{polarity_t}.$$

尽管亚马逊和百事分别属于电子商务和消费饮料这两个不同的行业，但该模型识别出了它们之间的一种微妙联系。纳入 Amazon_polarity_t 表明，对亚马逊的积极情绪可能反映出更广泛的消费者乐观情绪或消费能力，而这反过来又会对百事等消费品公司产生下游影响。这种关系并非显而易见，这也体现了大语言模型

尽管这样的阿尔法本身并不能保证具有强大的预测能力，但它在扩大候选阿尔法的范围方面发挥着重要作用。

4.2 股票预测性能

根据表 5，在大多数情况下，Transformer 在均方误差（MSE）方面表现最佳。Transformer 的绘图（图 4）显示，预测的股票价格与实际股票价格趋势密切吻合，这表明其捕捉数据中潜在模式的能力很强。

LSTM 的表现也不错，尽管不如 Transformer 准确。其预测通常能跟上实际价格走势，但偏差略大一些。同样，TCN 模型也能捕捉到股票价格趋势的大致形态。在某些时期，它的表现可以超过 LSTM。

对于传统机器学习模型而言，支持向量回归（SVR）的性能最差。随机森林集成方法有时能产生与 Transformer 相竞争的性能（例如，汇丰银行和丰田）。

总体而言，该分析突显了基于 Transformer 的架构在金融时间序列建模方面的优势，尤其是在有大型语言模型生成的信息性信号支持的情况下。此外，大语言模型生成的公式化特征具有很强的预测能力。我们将在下一章中进一步探讨这些结果。此外，图 3 展示了为丰田公司生成的阿尔法因子和推理过程的阿尔法因子。

此外，大语言模型在生成公式化阿尔法过程中的推理过程，可作为改进金融决策的宝贵参考。而且，大语言模型提供的新颖见解能启发人类投资者开发新的特征，以改进股票价格的预测。

表 4. 五家公司的阿尔法公式

时

蒙斯尼吉
泰古就是它
没什么 ss 或 s
The tindh anh asi& S eie S"

布（拟声词，

Aldoi

表 5. 丰田大语言模型生成的阿尔法因子进行股票价格预测的模型构方差对比

图 3：生成丰田阿尔法的推理过程

表 5. 丰田大语言模型生成的阿尔法因子进行股票价格预测的模型构方差对比

4.3 情感整合对预测准确性的影响

表 6 比较了 Transformer 模型在包含和不包含情感信息时的性能。每个股票都进行了十次实验。在这五家公司中，情感特征的融入显著提高了预测准确率，这体现在均方误差（MSE）的持续降低上。此外，在加入情感信息的设置中，标准差通常更小，表明模型表现更稳定、更稳健。这说明情感信息不仅有助于提升预测性能，还有助于减少不同运行之间的变异性。尽管不同公司的改进幅度有所差异，但这一一致的趋势凸显了

除了平均表现外，标准差结果还能让我们了解大语言模型生成的公式化阿尔法因子的稳定性。尽管每次训练都会涉及大语言模型新生成的一组阿尔法信号，但所有股票的标准差相对较低，这表明该模型能持续学习到有意义的模式。这意味着大语言模型生成的阿尔法因子具有很强的泛化能力，使得 Transformer 即使在输入条件变化的情况下也能保持稳定的表现。这种稳健性在金融应用中尤其有价值，因为在实际部署中，重复

5 进一步讨论

5.1 不同特征生成方法的比较

5.1.1 Featuretools 生成的特征

为了突出使用大语言模型（LLMs）进行特征生成的优势，我们对几种方法进行了比较。其中一种是

(c)

(d) LSTM

(e) 公共

(f) SVD

(g) 随机森林

图 4: 不同模型的预测结果对比

表 6: 包含和不包含情感信息的 Transformer 性能 (10 次运行的 MSE 均值和标准差)

对原始数据执行聚合操作。例如，它可以使用加法、减法或除法等基本数学运算来组合价格和情感数据。

表 7 展示了由 Featuretools 基于预测能力分数（PPS）[34] 生成的每家公司的五大特征。PPS 是一个介于 0 到 1 之间的分数，用于捕捉线性和非线性关系。对于每家公司，我们计算所有由 Featuretools 生成的阿尔法因子与收盘价之间的 PPS，并选出分数最高的五个阿尔法因子。这些选定的阿尔法因子随后被用作预测模型的输入特征。

表 7：由 Featuretools 生成的每家公司的五大特征（基于预测能力分数）

Feature	Apple	HSBC	Pepsi	Tencent	Toyota
Feature 1	Close - Google_polarity	ABSOLUTE(Close)	PERCENTILE(Close)	-(Close)	Close + Honda_polarity
Feature 2	Close - Intel_polarity	-(Close)	Close + Walmart_polarity	Apple_polarity - Close	BMW_polarity + Close
Feature 3	Close - TSLA_polarity	1 / Close	1 / Close	ABSOLUTE(Close)	BYD_polarity + Close
Feature 4	Close + Google_polarity	PERCENTILE(Close)	Close - Colgate_polarity	SQUARE_ROOT(Close)	Close - Ford_polarity
Feature 5	Close + Samsung_polarity	SQUARE_ROOT(Close)	-(Close)	Close + Tencent_polarity	Close - GM_polarity

根据表 7，该方法的缺点很明显。首先，生成的特征在数学上是有效的，但难以解释。例如，很难解释“Close - TSLA_polarity”在决策中的含义。此外，一些生成的特征存在冗余和噪声。例如，“Close + Google_polarity”和“Close - Google_polarity”具有高度相关性，这可能会使模型产生混淆并降低泛化能力。而且，生成的特征缺乏多样性，因为大多数公式都遵循相似的结构。

表 10 展示了 Featuretools 生成的阿尔法与大语言模型生成的阿尔法在股票预测性能上的对比。总体而言，与大语言模型生成的阿尔法相比，Featuretools 生成的阿尔法结果准确性更低。

5.1.2 人为定义的公式化阿尔法因子

另一种用于比较的方法是人为定义的公式化阿尔法因子。这些特征取自 Zura Kakushadze 的论文《101 个公式化阿尔法因子》[35]。这篇论文在量化金融领域声名远扬，在学术研究和行业实践中都被广泛引用。它收录了 101 个交易公式，这些公式是结合实际交易经验和历史市场行为开发出来的。这些阿尔法因子是在一家量化对冲基金中设计的，基于价格、成交量和其他市场变量。每个公式都是经过人工精心设计和测试的，旨在捕捉特定的市场模式或交易信号。由于这些阿尔法因子是公开可用且被广泛使用的，它们为与大语言模型生成的特征进行比较提供了一个强有力的标准化基准。

为了更有效地使用这些阿尔法因子，我们还应用了预测能力分数（PPS）来衡量每个阿尔法因子对股票收盘价的预测效果。

表 8 列出了每家公司从“101 个公式化阿尔法因子”中选出的排名前五的人为定义的公式化阿尔法因子。这些阿尔法因子对应的数学表达式详见表 9。如表所示，所选阿尔法因子均未包含情绪信息，而是主要依赖传统市场信号，例如均值回归（如阿尔法因子 5、24、57）、动量（如阿尔法因子 9、10、28）、波动率与相关性（如阿尔法因子 18、82）以及相对强度（如阿尔法因子 42）。此外，这些阿尔法因子的结构多样性相对有限。

表 10 展示了各种模型使用三种不同类型的阿尔法因子（大模型生成的、Featuretools 生成的和人工定义的）进行股票价格预测的结果。显然，在其他条件相同的情况下，使用人工定义的阿尔法因子作为输入时，预测性能会显著下降，相比之下，使用大模型生成的阿尔法因子和 Featuretools 生成的阿尔法因子时性能更好。这种性能差距凸显了利用大模型生成公式化阿尔法因子的优势，因为大模型生成的阿尔法因子可以通过整合最新信息动态适应股票的最新走势，而人工构建的阿尔法因子通常是静态的，容易出现阿尔法衰减，其预测能力会随着时间的推移逐渐减弱。

表 8：人类定义的五大大公式化阿尔法（基于预测能力得分）

Rank	Apple	HSBC	Pepsi	Tencent	Toyota
1	Alpha#18	Alpha#18	Alpha#18	Alpha#18	Alpha#18
2	Alpha#5	Alpha#42	Alpha#42	Alpha#42	Alpha#24
3	Alpha#42	Alpha#57	Alpha#24	Alpha#24	Alpha#9
4	Alpha#24	Alpha#28	Alpha#83	Alpha#83	Alpha#10
5	Alpha#9	Alpha#33	Alpha#9	Alpha#9	Alpha#42

表 9：与表 8 对应的阿尔法表达式

表 10. 不同阿尔法来源下股票价格预测模型的均方误差 (MSE) 对比

5.2 对大语言模型生成的阿尔法因子的进一步评估

除了传统的基于误差的评估指标（如均方误差），信息系数（IC）在金融领域常被用于评估收益预测（通常称为阿尔法信号）的预测质量 [36]。信息系数量化了预测收益与实际未来收益之间的相关性，有助于了解

IC 值的范围在 -1 到 $+1$ 之间。IC 值为 $+1$ 表示阿尔法信号与实际收益之间存在完美的正线性关系，这意味着阿尔法对资产的排名与它们的实际表现完全一致。IC 值为 -1 则意味着存在完美的反向关系，即

它们已实现回报的。IC 为 0 表明不存在线性相关性，这意味着阿尔法无法提供有意义的预测信息。

IC 通常使用皮尔逊相关系数计算，公式如下：

$$IC = \frac{Cov(\hat{r}_t, r_t)}{\sigma_{\hat{r}_t} \sigma_{r_t}} (2)$$

其中， \hat{r}_t 表示 t 时刻的预测收益， r_t 表示实际实现的收益， σ 表示标准差。

由于 IC 能够同时反映预测信号的强度和方向，因此它被广泛用于评估金融领域定量模型的性能。

表 11. 五个阿尔法（公司特定）的信息系数（IC）

五家不同公司的五个阿尔法值的 IC 值（表 11）显示出正负相关性的混合情况。重要的是，每家公司都有其独特的五个阿尔法值组合，这意味着苹果公司的阿尔法 1 与百事公司的阿尔法 1 并不相同，尽管它们的大多数阿尔法因子往往产生相对较低的信息系数值，这表明它们对股票回报的预测能力较弱或不一致。

总体而言，大语言模型生成的阿尔法因子对股票收益的预测能力较弱，但对股票价格的预测能力较强，这一点在表 5 中有所体现。这种差异可能源于以下事实：考虑到大语言模型处理大量相关数据的能力，股票价格的预测可能比收益的预测更具挑战性，因为后者需要更复杂的模型来捕捉市场中的非线性关系。

6 结论

本研究提出了一个新颖的框架，该框架将大语言模型生成的公式化特征与 Transformer 模型相结合，用于预测股票走势。公式化阿尔法因子源自基础输入，如股票特征、技术指标和情绪得分。在情绪数据方面，既考虑了目标公司，也考虑了其相关公司。相关公司最初是通过命名实体识别（NER）提取的，依据是它们在有关目标公司的新闻文章中的出现频率。大语言模型进一步筛选出与股票走势相关性强的最相关公司，并将其纳入公式化阿尔法因子中。为了进行性能比较，使用了包括 LSTM、TCN、SVR 和随机森林在内的模型。在未来的研究中，可以探索大语言模型生成的公式化阿尔法的数量会如何影响预测性能。可以使用不同的提

参考文献

- [1] 胡晓康。基于时间融合变压器的股价预测. 2021 年第三届国际机器学习大会，大数据与商业智能（MLBDBI），
- [2] 谢莉、陈正明、余胜。用于股票走势预测的深度卷积 Transformer 网络。《电子学》，第 13 卷（第 21 期），
- [3] 王淑珍。一种基于双向长短期记忆网络和改进 Transformer 的股票价格预测方法。《IEEE 接入》，
- [4] 朱利安·佩纳斯。理解阿尔法衰减。《管理科学》，68（5）：3966 - 3973，2022。

- [5] 唐子仪、陈子川、杨家瑞、麦家耀、郑永森、王可泽、陈锦瑞和林良。《AlphaAgent：借助大语言模型驱动的阿
- [6] 陈启昭、川岛博明。一种基于情感相关性的双 Transformer 模型股票价格预测新方法。预印本，Research
- [7] 阿德比伊·A·阿里约、阿德乌米·O·阿德乌米和查尔斯·K·阿约。使用 ARIMA 模型预测股票价格。见《2014
- [8] 约翰·卡姆韦勒·穆廷达和阿莫斯·基普科里尔·兰加特。使用组合 GARCH-AI 模型预测股票价格。《非洲科
- [9] 普图·钱德拉·普尔纳马、妮·卢赫·维维克·斯里·拉哈尤·吉南特拉、瓦扬·阿古斯·苏里亚·达尔马和普图·阿古斯·埃卡·达尔马·乌达亚纳。用于股票价格预测的支持向量机（SVM）与线性回归
- [10] 胡珍、朱杰和谢健。基于支持向量机的股票市场预测。见《2013 年第六届信息管理、创新管理与工业工程国际会议》，第 2 卷，第 115-118 页，2013 年。
- [11] 杜世培、郝德宏、李晓。基于随机森林的股票预测研究。见《2022 年 IEEE 第二届数据科学与计算机应用国际会
- [12] S. 尤瓦拉杰、库马兰·J. 钱尼、拉贾·R. 文斯顿和 G. 贾扬蒂。基于梯度提升机和技术指标的股票价格预测。见《2025 年智能与创新技术在计算、电气和电子领域国际会议（IITCEE）论文集》，第 1-6 页，2025 年。
- [13] 窦伟。基于 LSTM 神经网络的股票价格预测。见《2019 年人工智能与先进制造国际会议（AIAM）论文集》，第
- [14] 萨加尔·辛德、拉利特库马尔·瓦德瓦、奈尼什·莫汉、维沙尔·帕加尔、尼廷·谢尔杰和索汉·马内。基于 LSTM 的股票价格预测。见《2023 年第七届计算、通信、控制与自动化国际会议（ICCUBEA）》，第 1-7 页，2023
- [15] 阿迪勒·莫加尔和姆哈迈德·哈米什。使用长短期记忆循环神经网络进行股市预测。《计算机科学汇刊》，170:1168 - 1173, 2020。第 11 届国际环境系统、网络与技术会议（ANT）/ 第 3 届国际新兴数据与
- [16] 陈启昭、川岛弘明。基于大语言模型的 sentiment 分析在股票价格预测中的应用。见《2024 年 IEEE 大数据国际
- [17] 李硕哲、Zachery B Schulwol 和里斯托·米库莱宁。基于 Transformer 的股票时间序列预测，2025 年。
- [18] 哈尔西姆拉特·凯莱、叶乔和纳德·巴盖扎德。使用 Transformer 和情感分析支持股票趋势预测，2023 年。
- [19] 曾珍、拉奇尼特·考尔、苏切莎·西达甘加帕、萨巴·拉希米、塔克·鲍尔奇和曼努埃拉·贝洛索。使用 CNN 和
- [20] 李颖、王向荣、郭艳辉。CNN-Trans-SPP：一种用于股票价格预测的结合 CNN 的小型 Transformer。《电子研究
- [21] 亚什·K·帕德希与普丽蒂·凯尔教授。《基于机器学习的股票市场技术分析指标：比较分析》。发表于《2021 年第 12 届计算通信与网络技术国际会议（ICCCNT）》，第 1-6 页，2021 年。
- [22] 拉马赞·法鲁克·奥古兹、亚辛·乌伊贡、穆罕默德·S. 阿克塔什和伊沙克·艾库尔特。关于使用技术分析指标预测股票市场价格变动方向。见《2019 年第 27 届信号处理与通信应用会议（SIU）》，第 1-4 页，2019 年。
- [23] 阿卡什·迪普、阿布塔利布·希尔瓦尼、克里斯·莫尼科、斯维特洛扎尔·拉切夫和弗兰克·J. 法博齐。高频交
- [24] 约翰·潘与张洪福。《利用基本面分析进行股票趋势预测以获利》，2024 年。
- [25] K. Puneeth、Rudagi Sagar、M. Namratha、Patil Ranispoorti 和 Wadi Rohini。比较研究：使用基本面分析和

- [26] 莉拉·萨蒂什·帕尔瓦萨、德武·纳加·维拉·塔伦、门特姆·叶斯万斯和乔纳拉加达·苏里亚·基兰。《基于情感分析和增量聚类方法的股票市场预测》。发表于《2023 年第 9 届高级计算与通信系统国际会议 (ICACCS)》，第
- [27] 尤金·F·法玛和肯尼斯·R·弗伦奇。股票和债券收益中的共同风险因素。《金融经济学杂志》，33
- [28] 任伟哲、秦怡辰和李阳。《通过热启动遗传编程进行量化投资的阿尔法挖掘与增强》，2024 年。
- [29] 赵俊杰、张承熙、秦敏、杨鹏。Quantfactor 强化学习：利用方差有界强化学习挖掘稳定的公式化阿尔法因子，
- [30] 于硕、薛红岩、敖翔、潘飞扬、何佳、涂丹丹、何清。通过强化学习生成协同的公式化阿尔法集合。《第 29 届 ACM SIGKDD 知识发现与数据挖掘会议论文集》(KDD '23)，第 5476–5486 页。美国计算机协会，2023 年 8 月。
- [31] 陈启昭。基于提示词的大语言模型结合强化学习增强的事后调整预测股价变动。见《2025 年第四届大数据、区块链与经济管理国际会议论文集》(ICBBEM 2025)，第 475–483 页。亚特兰蒂斯出版社，2025
- [32] C. Hutto 和 Eric Gilbert。Vader：一种用于社交媒体文本情感分析的简洁规则模型。《国际人工智能协会网络与社交媒体会议论文集》(ICWSM)，第 216–225 页。2009 年。
- [33] 詹姆斯·马克斯·坎特和卡利安·维拉马查尼尼。深度特征合成：迈向自动化数据科学探索。见《2015 年 IEEE 数据科学与高级分析国际会议论文集》，DSAA 2015，法国巴黎，2015 年 10 月 19–21 日，第 1–10 页。IEEE，2015
- [34] 弗洛里安·韦奇奥雷克、托比亚斯·克拉贝尔和苏里亚·克里希纳穆尔蒂。8080labs/ppscore：Zenodo 发布，
- [35] 祖拉·卡库沙泽。101 个公式化阿尔法因子，
- [36] 张峰、郭瑞特和曹宏高。信息系数作为股票选择模型的绩效衡量指标，2020 年。