

Alpha 掘金系列之五

金融工程专题报告

证券研究报告

金融工程组

分析师：高智威（执业 S1130522110003） 联系人：王小康

gaozhiw@gjzq.com.cn

wangxiaokang@gjzq.com.cn

如何利用 ChatGPT 挖掘高频选股因子？

ChatGPT 模型介绍及原理解析

GPT (Generative Pre-trained Transformer) 是一种大语言模型 (LLM)，能够学习大量文本数据，并推断出文本中词语之间的关系。ChatGPT 能够进行连续对话，综合上下文内容进行交流，能完成翻译、撰写邮件、代码等任务。该模型相较于传统 LSTM 模型的改进之处在于其引用了 Transformer 模型，对输入数据的不同部分给予不同权重。

ChatGPT 之所以能够获得如此高的智能水平，参数数量提升所带来的涌现现象 (Scaling Law) 和加入 RLHF (人类反馈的强化学习) 所带来的对于人类偏好理解的提升起到了重要作用。RLHF 训练共分为 3 步，首先聘请 40 名标注员对指令进行标注，对模型进行微调。然后对模型的不同输出结果进行排序，使其更符合人类预期，并利用排序结果训练一个打分模型 (Reward Model)。最终采样新的指令作为输入数据，根据打分模型进一步优化模型的输出结果。结合打分模型训练，得到最终的 ChatGPT 模型。

ChatGPT 提示工程介绍及使用指南

提示工程 (Prompt Engineering) 主要用于开发和优化语言模型中的提示，有效地将 ChatGPT 用于各种应用和研究主题。掌握并应用好提示工程的技能，不仅能够提高使用人工智能系统的准确性和效率，也能够降低成本并提升使用体验。最基本的提示公式包括角色、任务及指令三个部分，其主要目的在于使模型对于所需要的回答类型和回答方式有一定的指向性。提示的内容越详尽、精确，模型能够给出的回答更能符合我们的预期，从而更便捷得到我们需要的结果。除标准的提示公式外，也有多种针对不同任务类型的提示方式。一种独特的提示方式为思维链提示 (Chain-of-Thought Prompting)，其主要思路为将一个复杂问题拆分成多个步骤，引导模型逐步思考并进行纠偏，最终得到需要的结果。

ChatGPT 因子挖掘实战

ChatGPT 在量化研究领域同样拥有广泛的使用前景，我们以最常见的因子挖掘作为测试场景，考察模型经过一定的提示后，能否给出符合需求的结果。在中低频领域，ChatGPT 给出了价和量的变异系数因子，我们利用 5 日滚动数据构建因子并进行周度调仓测试。发现因子 IC 指标表现较好，但多头组超额收益较低，难以成功构建投资策略。

另外，我们令 ChatGPT 模型尝试利用高频数据构建出独特因子，并限定其数据使用范围为委托价和委托量。模型经过一定指导后给出了买卖盘力量因子，经过测试发现买卖盘力量差异因子在日频上表现优异，多头年化超额收益率达到 17.29%，但因子衰减速度较快。为符合交易实际，我们针对买盘和卖盘力量因子分别降至周频进行测试，发现虽然因子整体多空单调性一般，但多头组合表现尚可，多头年化超额收益率分别为 9.77% 和 10.20%。最终，我们利用相对表现较好的卖盘力量因子构建中证 1000 指数增强策略。发现在单边千分之二的手续费率下，策略的年化超额收益率为 7.17%，信息比率为 0.57。

此外我们对于 ChatGPT 代码能力进行测试，发现针对常用的量化研究所需框架、函数等能较准确的给出结果，但使用时需要注意代码细节，确保其符合实际需求。对模型所给代码进行微调可以大幅提升研究效率。

风险提示

- 1、ChatGPT 模型具有一定的随机性，在部分情况下可能回答错误，不符合用户需求与认知，并影响到用户判断。
- 2、以上因子测试结果通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。
- 3、策略依据一定的假设通过历史数据回测得到，当交易成本提高或其他条件改变时，可能导致策略收益下降甚至出现亏损。

内容目录

一、ChatGPT 模型介绍及原理解析	4
1.1 大语言模型与模型演进历程	4
1.2 ChatGPT 模型表现优异背后的逻辑	5
二、ChatGPT 提示工程介绍及使用指南	5
2.1 Prompt 提示公式	6
2.2 提示工程分类	7
2.3 思维链提示 (Chain-of-Thought Prompting)	8
三、ChatGPT 因子挖掘实战	10
3.1 周频变异系数因子构建与测试结果	10
3.2 高频买卖盘力量因子构建与测试结果	12
3.3 ChatGPT 代码输出测试：因子测试与组合优化	15
四、总结	17
风险提示	18

图表目录

图表 1： ChatGPT 模型演进历程	4
图表 2： RLHF (人类反馈的强化学习) 过程示意图	5
图表 3： ChatGPT 生成因子示例 1	6
图表 4： ChatGPT 生成因子示例 2	7
图表 5： ChatGPT 生成因子示例 3	7
图表 6： 提示工程分类示例及说明	7
图表 7： ChatGPT 情感分析示例 1	8
图表 8： ChatGPT 情感分析示例 2	8
图表 9： A 因子 IC 指标示例 1	8
图表 10： B 因子 IC 指标示例 2	8
图表 11： ChatGPT 思维链提示示例 1	9
图表 12： ChatGPT 思维链提示示例 2	9
图表 13： ChatGPT 思维链提示示例 3	9
图表 14： ChatGPT 思维链提示示例 4	10
图表 15： ChatGPT 思维链提示示例 5	10
图表 16： ChatGPT 生成因子示例 4	10

图表 17: ChatGPT 变异系数因子 IC 指标 (周频)	11
图表 18: ChatGPT 变异系数因子多空组合净值 (周频)	11
图表 19: ChatGPT 变异系数因子多空组合指标 (周频)	11
图表 20: ChatGPT 高频因子挖掘示例 1	12
图表 21: ChatGPT 高频因子挖掘示例 2	12
图表 22: ChatGPT 买卖盘力量因子 IC 指标 (日频)	13
图表 23: ChatGPT 买卖盘力量因子多空组合净值 (日频)	13
图表 24: ChatGPT 买卖盘力量因子多空组合指标 (日频)	13
图表 25: ChatGPT 买卖盘力量因子多头组合净值 (周频)	14
图表 26: ChatGPT 买卖盘力量因子多空组合指标 (周频)	14
图表 27: 基于 ChatGPT 买卖盘力量因子的中证 1000 指数增强策略表现	15
图表 28: 基于 ChatGPT 买卖盘力量因子的中证 1000 指数增强策略指标	15
图表 29: ChatGPT 因子测试框架代码示例 1	16
图表 30: ChatGPT 因子测试框架代码示例 2	16
图表 31: ChatGPT 组合优化代码示例 1	17
图表 32: ChatGPT 组合优化代码示例 2	17

一、ChatGPT 模型介绍及原理解析

在前期系列报告中，我们利用高频数据已经构建出了一系列表现优异的选股因子。本篇报告作为 Alpha 掘金系列的第五篇，同时也是 ChatGPT 量化研究系列的第一篇，将使用 ChatGPT 探索其在量化研究领域的表现。经过测试，发现模型经过一定的引导能够挖掘出具有一定创新性的高频因子，最终得到了 ChatGPT 买卖盘力量因子。在将因子降至周频后，成功构建出了能够满足机构投资者要求的中证 1000 指数增强策略。

1.1 大语言模型与模型演进历程

ChatGPT（全名：Chat Generative Pre-trained Transformer）是由人工智能研究实验室 OpenAI 在 2022 年 11 月 30 日发布的一款 AI 聊天程序，一经发布便在全球开启了一波 ChatGPT 热潮，短短 2 个月内用户数已过亿。该模型是一款人工智能技术驱动的自然语言处理工具，使用的是 GPT-3.5 架构的千亿参数大语言，能够进行连续对话、综合上下文内容进行交流的自然语言处理（NLP）模型，并通过强化学习进行训练，从而理解人类的语言来进行对话，甚至能完成撰写邮件、视频脚本、文案、翻译、代码，写论文等任务。ChatGPT 也是迄今为止 AI 大模型最接近商用落地的成果，输出内容非常接近人类的常识、认知、需求和价值观。这项新的科技革命正在持续升温，所覆盖的范围在新媒体、编程、教育、医疗、广告、电商平台等各行各业。然而针对二级市场投资研究领域，ChatGPT 模型能带来怎样的变革，相关研究较少，本文将从量化研究领域为读者带来该模型的使用方法和效果演示。

GPT（Generative Pre-trained Transformer）作为一种大语言模型（LLM），能够学习大量文本数据，并推断出文本中词语之间的关系。随着过去几年计算能力的不断发展，输入数据集和参数空间（parameter space）的不断增加，LLM 的能力也在不断提升。

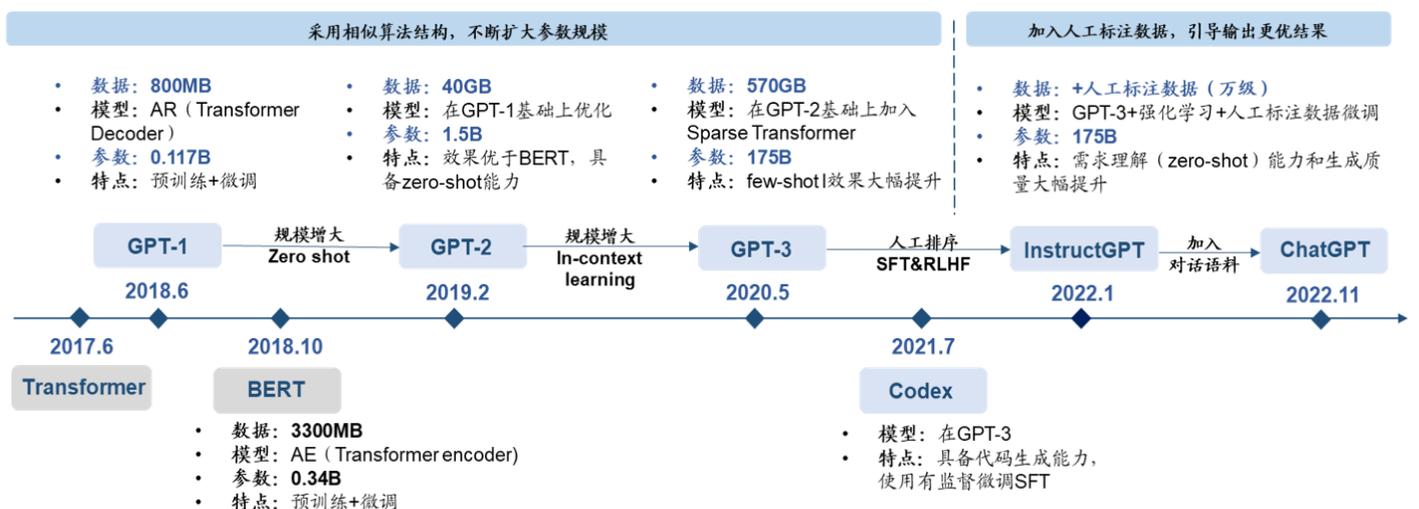
语言模型的一项基本任务是预测一句话中的单个词，或根据上文推断下文。传统的 LSTM（Long Short-Term Memory）模型在处理这类问题时存在两大缺陷：

- 模型对于文本中不同词无法区分出重要程度的差异，即无法让某些词比其他词的权重更高。
- 该模型对于长距离的信息不能有效提取和记忆，导致了大量信息丢失。

2017 年论文《Attention is All You Need》中 Transformer 的引入有效解决了上述问题。它通过跟踪序列数据中的关系来学习上下文并学习语句的含义，利用自注意力机制（Self-attention mechanism）给予输入数据的不同部分赋予不同权重，这一革新让 LLM 拥有了更大的成长空间，同时也能够处理更大的数据集。

2018 年 6 月发布的 GPT-1 模型就已经开始使用 Transformer 模型，该模型包括了编码器和解码器结构，参数数量大约为 1.17 亿个。2019 年 2 月 GPT-2 模型的参数数量已经达到 15 亿，在自然语言处理领域取得了非常显著的进展。GPT-3 模型的参数数量高达 1750 亿，成为当时最先进的自然语言处理领域的最先进模型。2023 年 3 月发布的 GPT-4 模型，将传统的文本输入拓展到了图片，同时它的理解能力也得到了显著提升，在处理复杂问题时表现出了更高的准确性和洞察力。

图表1: ChatGPT 模型演进历程



来源：《On the Comparability of Pre-trained Language Models》(Matthias, 2020)，《Evaluating Large Language Models Trained on Code》(Mark Chen, 2021)，OpenAI, BFT 智能机器人研究公众号，国金证券研究所

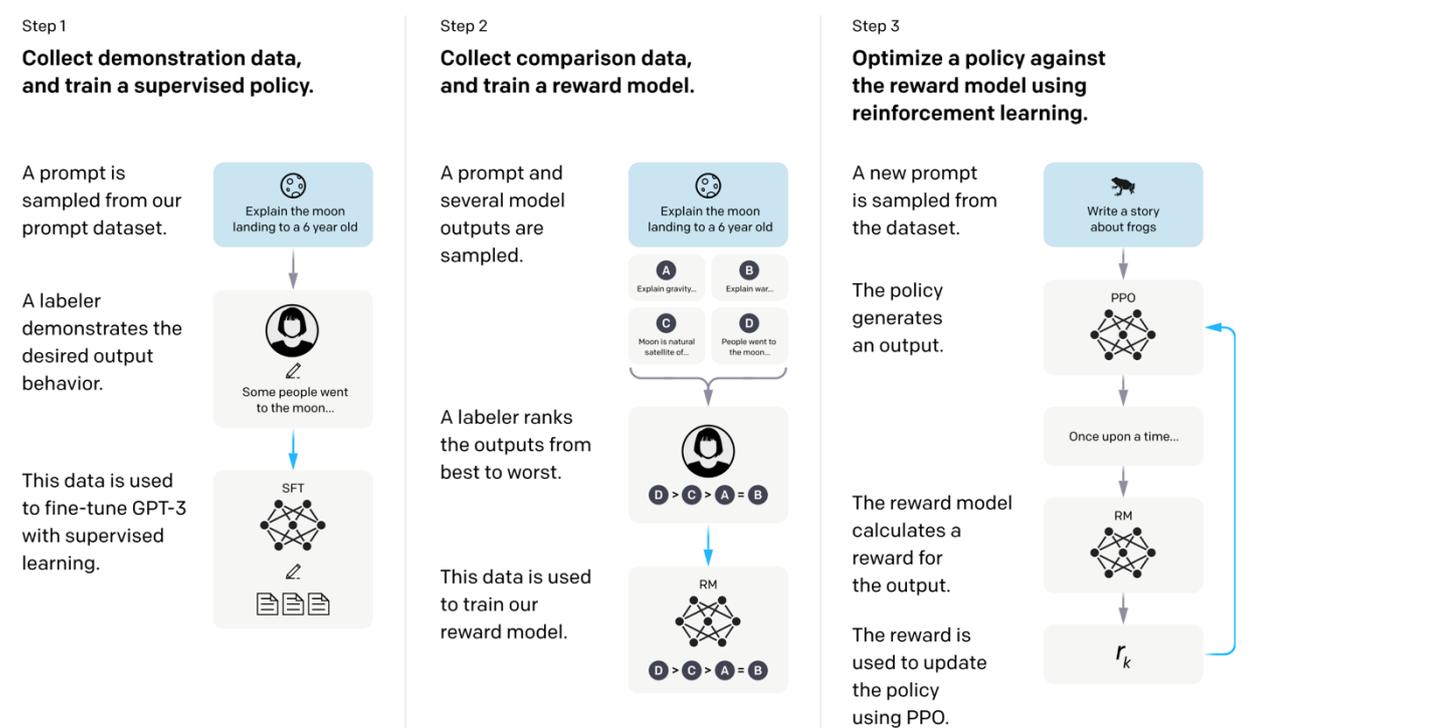
1.2 ChatGPT 模型表现优异背后的逻辑

ChatGPT 拥有如此高度的智能水平背后有着多重因素的共同作用，包括随着参数数量提升带来的涌现现象（scaling law）、加入 RLHF（人类反馈的强化学习）所带来的对于人类偏好理解的提升等都给模型提供了较强的提升作用。

RLHF 是一种用于增强语言模型性能的技术，利用关系网络和潜在因子来增强模型的代表能力和泛化能力。这种方式可以帮助其更好地理解和学习文本中的语义关系和结构，生成更加准确和流畅的文本。其训练基本步骤如下：

- 从大量的包含人类真实意图的指令集中采样，聘请 40 名标注人员对这些指令进行标注，标注完成后在 ChatGPT 中对模型进行微调。
- 再次从指令集中采样作为输入数据，从 ChatGPT3.5 中得到多个不同结果，聘请标注员对这些输出的好坏顺序进行标注。得到输出顺序后，训练一个打分模型（reward model），使得模型输出结果更能贴近人类最需要的答案。
- 获得打分模型后，从指令集中采样新的指令作为输入数据，根据打分模型进一步优化模型的输出结果。结合打分模型训练，得到最终的 ChatGPT 模型。

图表2: RLHF (人类反馈的强化学习) 过程示意图



来源: <https://openai.com/blog/chatgpt>, 国金证券研究所

在这种训练模式下，AI 逐步具备了一些“常识”。自 GPT-2 开始，只需给模型投喂一些示例，就能使其举一反三给出需要的结果。基本的训练模式分类包括：

- **Zero-shot learning** 指的是模型在没有经过训练数据的情况下，可以利用已有的知识和推理能力完成新任务。例如，在翻译任务中如果模型训练过英文到法语的翻译任务，它可以无需训练来处理英文到德语的翻译任务。
- **One-shot learning** 指的是模型在只有一组训练数据的情况下，可以学习并完成任务。例如，在人脸识别中，如果只提供一张某人的照片，模型可以学习识别该人的所有照片。
- **Few-shot learning** 指的是模型在只有少量训练数据的情况下，可以学习并完成任务。例如，在自然语言推理中，如果只提供很少的相关语料，模型可以学习理解和推理其他类似语料的能力。

因此，ChatGPT 由于巨大的参数数量和大量的预训练，拥有极强的学习能力，可以在不同的任务中做到足够出色的表现。

二、ChatGPT 提示工程介绍及使用指南

大语言模型能够处理的任务极其广泛，人类与其交流的灵活度和自由度也达到了空前的高度。然而如何正确与其对话、进行合适的提示对于获取最终需要的结果至关重要。提示工程（Prompt Engineering）作为一门新兴学科，主要用于开发和优化语言模型中的提示，有效地将 ChatGPT 用于各种应用和研究主题。掌握并应用好提示工程的技能，不仅能够提高使用人工智能系统的准确性和效率，也能够降低成本并提升使用体验。对于研究人员而言，也有助于更好地理解

解模型的能力和局限性，通过交互和提示，更加方便准确地使 ChatGPT 给出我们需要的结果。

2.1 Prompt 提示公式

提示工程的核心是通过设计和提供有效提示或起始短语，指导 ChatGPT 语言模型的文本生成。一般的提示工程通常由以下 3 个元素组成：

- 角色：模型在本文生成任务中应扮演的角色。
- 任务：对模型需要生成的内容进行清晰简介的叙述。
- 指令：在生成文本时应遵循的指令。

提示公式一般形式为：“作为一个【角色】，按照这些指示生成【任务】：【指令】”。如希望 ChatGPT 充当客服的角色完成日常的客户答疑等场景需求，则提示公式可以为：

- 作为一名【客服人员】，请生成【客户询问的回复】：【这些回复应当专业、准确、简洁】。

限制提示方式的一个重要作用在于，这种方式能够向模型灌输一种“角色”，并且起到“约束行为”的作用，从而能使模型按照人类的期望完成回答。

此外，我们可以为模型设置更具体化的角色以满足我们的要求。如我们需要模型能给出一个足够专业的回答时，可以针对【角色】部分稍加修饰，以使模型的回复更具针对性。针对【指令】部分，我们也可以完善、丰富提示信息，进一步限制模型的回答范围，从而更便捷得到我们需要的结果。如：

- 作为一名【资深的量化研究员】，请给出一个你认为行之有效的选股因子。【该因子必须：具有原创性、具有较强的逻辑支撑、且因子也应当有出色的业绩表现】。

图表3: ChatGPT 生成因子示例 1

 作为一名资深的量化研究员，请给出一个你认为行之有效的选股因子。该因子必须：具有原创性、具有较强的逻辑支撑、且因子也应当有出色的业绩表现。

 作为一名量化研究员，我会选择一个基于机器学习技术的原创选股因子：利用自然语言处理技术从公司年报中提取关键词，通过计算关键词的情感得分来评估公司的财务健康和发展前景。

这个因子的逻辑支撑是，公司年报中的文字材料可以提供有关公司业务和经营环境的丰富信息，而这些信息可以用来预测公司未来的业绩表现。通过使用自然语言处理技术来分析公司年报，我们可以快速而准确地提取出关键词，并计算这些词的情感得分。这些得分可以用来评估公司的整体情绪和态度，以及市场前景和风险。

这个选股因子的业绩表现也很出色。在我们的回测中，使用这个因子进行选股的组合表现优于市场平均水平，并取得了可观的超额收益。此外，我们还可以将这个因子与其他因子组合使用，以进一步提高投资组合的表现。

总之，利用自然语言处理技术从公司年报中提取关键词，并通过计算情感得分来评估公司的财务健康和发展前景，是一个具有原创性、逻辑支撑强、并且有出色业绩表现的选股因子。

来源：ChatGPT，国金证券研究所

以此为基础，我们可以进一步限定因子类别、构建方式和使用数据，最终得到一个符合我们基本需要的因子。在下图示例中，我们将提示的【指令部分】进行拓展，针对其回答中可能存在的偏差进行限制，从而大大提升了因子的有效性。

图表4: ChatGPT 生成因子示例 2

作为一名资深的量化研究员，请构造出一个你认为行之有效的选股因子，并给出详细的构建方式。该因子必须：

1. 具有原创性
2. 具有较强的逻辑支撑
3. 使用逐笔成交数据中的日内高频量价数据构建
4. 利用高频数据降频形成日频或周频因子

我可以提供一个基于逐笔成交数据的选股因子，称为“价格漂移率因子 (Price Drift Factor)”。

1. 原创性：该因子的构建是基于市场价格的时间序列漂移特征，结合量化交易中价格漂移 (Price Drift) 理论得出的。
2. 逻辑支撑：根据价格漂移理论，市场中的价格是受到市场交易参与者的心理预期和实际交易行为的影响。当市场价格下跌时，交易参与者可能会在价格触及心理支撑位时出现买盘，从而支撑市场价格上涨；当市场价格上涨时，交易参与者可能会在价格触及心理压力位时出现卖盘，从而导致市场价格下跌。因此，可以通过计算市场价格在某一时刻前后的漂移情况，来判断市场交易者的心理预期和实际交易行为，并用此构建选股因子。
3. 业绩表现：我们可以利用该因子在历史数据上的表现来评估其有效性。我们可以回测过去一段时间内使用该因子的选股策略，并与市场基准指数进行比较。如果该因子的选股策略在历史数据上表现出色，那么该因子就可以被认为是有效的。
4. 构建方式：该因子的构建方式如下：
 - 首先，我们需要获取逐笔成交数据。
 - 对于每一笔成交，记录成交价和成交量。
 - 对于每个时间窗口（如5分钟），计算这段时间内市场价格的漂移率。漂移率的计算公式为：

$$\text{漂移率} = (\text{当前价格} - \text{前}n\text{个时间窗口的平均价格}) / \text{前}n\text{个时间窗口的平均价格}$$
 其中， n 是一个参数，可以根据实际情况进行调整。
 - 最后，对于每个股票，计算其在过去一段时间内漂移率的平均值，并将其作为选股因子的值。
5. 降频方式：由于该因子使用的是高频数据，我们需要将其降频形成日频或周频因子。我们可以在日内每隔一段时间计算漂移率，并将其记录下来。然后，在每个交易日结束后，取当天所有漂移率的平均值作为该股票的当日因子值；在每个交易周结束后，取本周所有交易日的漂移率平均值作为该股票的周度因子值。

来源：ChatGPT，国金证券研究所

图表5: ChatGPT 生成因子示例 3

作为一名资深量化研究员，按照以下指令，构建一个选股因子：

1. 不要使用现有的技术指标去直接构建
2. 构建一个新的、独特的因子
3. 因子的构建逻辑不要太简单
4. 考虑使用价量类数据去研究流动性
5. 请给出构建的详细步骤和计算方法

作为一名资深量化研究员，我将按照以下指令构建一个新的、独特的选股因子，并使用价量类数据去研究流动性。以下是我的详细步骤和计算方法：

1. 我将选择使用股票的价格和成交量数据，因为这是股票市场中最基本、最重要的数据之一。价格和成交量通常可以反映股票的流动性和市场情绪，也是很多交易策略的基础。
2. 我将计算每个交易日的股票价格的滚动平均值和滚动标准差，并分别表示为MA和STD。同时，我也将计算每个交易日的成交量的滚动平均值和滚动标准差，并分别表示为VMA和VSTD。
3. 接下来，我将计算股票价格的流动性指标LI (Liquidity Index)，它是MA和STD之间的比率，用于衡量价格的波动性和市场流动性。LI的计算公式如下：

$$LI = STD / MA$$
4. 我将计算成交量的流动性指标VLI (Volume Liquidity Index)，它是VMA和VSTD之间的比率，用于衡量成交量的波动性和市场流动性。VLI的计算公式如下：

$$VLI = VSTD / VMA$$
5. 我将使用LI和VLI两个流动性指标来构建新的选股因子，称为“流动性因子 (LF)”。LF的计算公式如下：

$$LF = LI * VLI$$
6. 最后，我将对所有股票进行排名，选取LF值最高的股票作为买入信号，选取LF值最低的股票作为卖出信号。

总体来说，LF因子通过考虑股票价格和成交量的流动性指标，综合反映了股票市场的情况，可以作为一种有效的选股因子，用于寻找市场流动性较好的股票。

来源：ChatGPT，国金证券研究所

可以看出，经过详细的、严格的限制后，模型已经能给出一些基本符合要求的因子。在示例 2 中，模型不仅正确使用了逐笔成交数据中的字段，且根据提示中的限制条件，给出了使用计算均值的方法将因子进行降频的操作。示例 3 中，因子的构建方式可以明显看出模型参考了一些现有文献和研究结果，在其已有知识的基础上进行一定变形、组合得到了新的选股因子。

2.2 提示工程分类

除标准化的提示工程之外，我们可以使用多种类型的提示方式，使模型完成不同类型任务，满足用户的各类需求。

图表6: 提示工程分类示例及说明

提示类型	提示示例	提示类型说明
知识生成提示	将一些【新信息】与关于【特定主题】的现有知识结合，生成新的知识。	使用模型里预先存在的知识，生成新的信息或回答问题。
情感分析提示	请对以下产品评论进行情感分析，并将其分类为积极、消极或中立【插入评论】。	让模型判断一段输入文本的情感色彩或态度。
强化学习提示	使用强化学习将以下文本【文本】从【插入语言】翻译为【插入语言】。	通过向模型提供一组输入和奖励，并允许其根据所接受的奖励调整行为，从而让模型从过去的行为中学习。
对抗性提示	生成难以被分类为具有【插入情感】情感的文本。	让模型生成的文本难以生成与所需一致的文本。

来源：国金证券研究所

针对其中情感分析类型，我们截取了两段不同的新闻热点，以供模型识别。发现由于神经网络模型本身具有一定的随机性，当输入文本没有特别明确的情感词时，模型给出的结果会有出一定的不确定性。但整体而言，仍能保证一定的准确率。

图表7: ChatGPT 情感分析示例 1

WX 请给以下文本做出情感分析，并分类为积极、消极或中立：
【父亲突发疾病去世，“90后”接管400亿上市公司，创始人遗嘱现身股东大会，质疑选举违规】3月23日，宁波杉杉股份有限公司召开2023年第一次临时股东大会、第十届董事会第四十次会议，选举郑永刚为公司第十届董事会董事长。据《国际金融报》24日报道，有知情人士透露，当天郑永刚遗嘱周婷出现在会议现场，并指称该次股东大会是违规和错误的。

据知情人士透露，3月23日下午，周婷出现在会场，作为已故杉杉股份实际控制人、董事长兼法定代表人郑永刚的现任配偶及3名亲生子女的法定监护人，她向到场股东表示，基于继承关系，她应当成为杉杉股份的实际控制人。

知情人称，周婷表示，董事会擅自审议未经自己审阅和同意的议案并对外发布，是违规和错误的。她认为，董事会的做法对她和子女应该合法继承的财产及权利造成了损害，也违背了郑永刚先生的遗愿；并且，上市公司治理结构与实控人完全脱节，可能对杉杉股份治理结构、规范运作产生重大不利影响，继而引发上市公司合规性风险。

情感分析： 这篇文本的情感倾向为消极，因为它描述了一系列问题和争议，包括杉杉股份的治理结构和选举程序的争议，以及已故创始人的遗愿被违反所引起的财产和权利损害。文本中没有体现积极的元素，也没有表达中立的观点。

图表8: ChatGPT 情感分析示例 2

WX 请给以下文本做出情感分析，并分类为积极、消极或中立：
2022年，中国石化油气勘探开发业务盈利创近十年最好水平，境内油气储量替代率165%，全年油气当量产量6878.96万吨，同比增长1.9%，境内原油产量3532.26万吨，天然气产量353.65亿立方米，同比增长4.1%，境内油气当量产量创历史新高，盈利创近十年最好水平。

情感分析： 积极

该文本描述了中国石化油气勘探开发业务在2022年的业绩表现，其中油气储量替代率达到165%，全年油气当量产量、原油产量和天然气产量均有所增长，油气当量产量更是创历史新高。最后，该公司的盈利也达到近十年最好水平。因此，该文本表现出了积极的情感色彩。

投资建议：由于该公司在2022年的业绩表现较好，因此该公司可能是一个值得投资的公司，然而，投资涉及到众多因素，还需要进一步的分析。

来源：每日经济新闻，ChatGPT，国金证券研究所

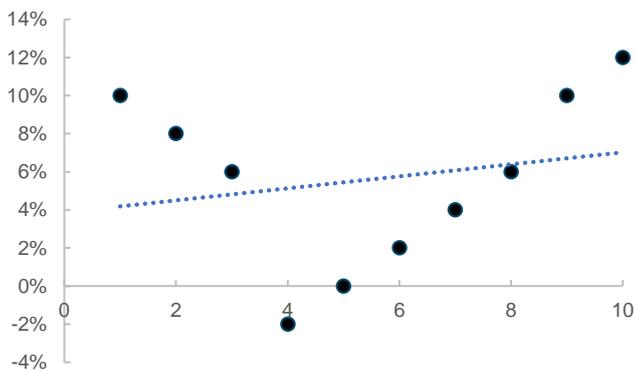
来源：e 公司公众号，ChatGPT，国金证券研究所

2.3 思维链提示 (Chain-of-Thought Prompting)

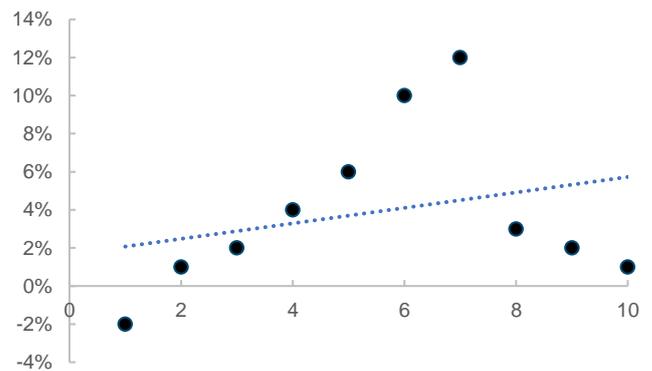
Wei et al (2022) 发现利用思维链 (Chain of thought) 能够极大提升大语言模型在处理复杂逻辑问题上的表现。通过简单的思维链提示的方式就能大大提升一些通用任务、符号化的推理任务和算数任务的准确率。这种提示方式一般采用分步骤思考的方式 (think step by step) 进行，引导模型将中间推导过程展现出来。一般而言，思维链提示更多用于数学推理提升的演示，此处我们提供示例说明思维链提示的方式也能让模型对于量化研究领域提升其回答效果。

我们知道，IC 指标 (秩相关系数) 作为一种因子评价方式被广泛使用，仍存在一些明显的漏洞。对于两个相同 IC 值得因子，若其因子值与收益率的相关性在不同区间范围内有着明显的差异，可能会导致使用该因子在投资时实际效果会大打折扣。如下图所示，我们给出了两个因子值排序和对应收益率的一一对应关系。两因子的 IC 值分别为：0.22 和 0.25。然而由图可知，A 因子虽然 IC 值相对较低，但多头表现较好。而 B 因子虽 IC 值相对较高但多头组却表现最差。这一现象会使得根据因子进行实际投资时得到与 IC 指标截然相反的效果。

图表9: A 因子 IC 指标示例 1



图表10: B 因子 IC 指标示例 2



来源：国金证券研究所

来源：国金证券研究所

一般而言，根据分组赋予 IC 不同权重，多头组权重更高，求得加权 IC 是一个比较直接的解决方案。接下来我们利用思维链提示的方式引导 ChatGPT 模型思考，判断其能否准确意识到这一问题。并根据此问题得出改进解决方案。

图表11: ChatGPT 思维链提示示例 1

WX IC作为因子评价的一种指标存在一些缺点，请问有什么方式可以修改IC

IC指标衡量了一个因子在样本期内的预测能力，但是它也存在一些缺点，例如可能会受到极端值、样本选取、市场环境等因素的影响，因此需要一些方法来修正或优化。

以下是一些修改IC指标的方法：

1. Spearman's Rank Correlation Coefficient: 与Pearson相关系数不同，Spearman的相关性基于等级，而不是实际值。使用Spearman相关系数可以减少因为极端值对IC的影响。
2. IC-adjusted Return: 将IC指标与因子收益率结合使用可以提高因子评价的稳定性和准确性。例如，使用IC值来加权因子收益率，以更好地衡量因子的预测能力。
3. IC Rank-T Test: 这是一种比较IC指标在不同时间段内是否有显著差异的方法。通过进行假设检验，我们可以判断IC值的变化是否足以说明因子的预测能力在时间上的显著性差异。
4. IC Decay Rate: 对于长期持有的投资者来说，使用长期IC值可能更具有参考价值。使用IC衰减速率，可以对IC值进行加权，以更好地衡量因子的长期预测能力。
5. Industry-adjusted IC: 如果因子在特定行业中的预测能力较强，则可以使用行业加权IC来更好地衡量其绩效。

这些方法可以根据需要进行适当的组合使用，以改进IC指标的评价。同时，在使用任何因子评估方法之前，应该对因子本身进行全面的分析和验证，以确保其有效性。

来源：ChatGPT，国金证券研究所

我们首先尝试了直接询问 ChatGPT，有什么方式可以改进 IC 指标，发现其回答在不经适当引导的情况下很难给出我们思考的问题和改进角度。

图表12: ChatGPT 思维链提示示例 2

WX 请问你知道IC作为选股因子的评价指标，都有哪些缺点吗

IC (Information Coefficient) 是衡量某个选股因子与股票收益率之间相关性强度的指标。虽然 IC 是一种常用的指标，但也存在以下一些缺点：

1. 对极值敏感：当选股因子中出现极值时，IC 可能会被极值所主导，从而无法准确反映选股因子的实际效果。
2. 不考虑交易成本：选股因子在实际交易中可能存在交易成本，但 IC 并没有考虑这一点，从而无法全面评估选股因子的实际效果。
3. 不考虑实时性：选股因子的有效性可能会随着时间的推移而发生变化，但 IC 并没有考虑这一点，从而无法全面评估选股因子的实际效果。
4. 受限于样本选择：选股因子的有效性可能受到样本选择的影响，如果样本选择不合理，IC 将无法准确反映选股因子的实际效果。
5. 不考虑投资组合效应：选股因子的有效性可能会因为与其他选股因子的组合方式而发生变化，但 IC 并没有考虑这一点，从而无法全面评估选股因子的实际效果。

综上所述，虽然 IC 是一种常用的选股因子评价指标，但在实际应用中需要注意以上缺点，以免出现评估误差。

来源：ChatGPT，国金证券研究所

图表13: ChatGPT 思维链提示示例 3

WX 对于两个相同IC值的因子A和B，若都按因子值划分为10组，A因子多头组的因子值与收益率相关性较高，但B因子在多头组的因子值与收益率相关性较低。结果会导致两个因子在最终的收益表现上有较大差异。请问这一问题应通过什么方式修改IC指标进行改进。

针对上述情况，可以使用“横截面IC”（Cross-sectional IC）来修正IC指标，以更准确地反映因子与收益率之间的相关性。

横截面IC是指将选股因子分为多组后，在每组内计算因子值与收益率的相关系数，然后将所有组内相关系数的平均值作为该选股因子的横截面IC值。相对于传统的IC指标，横截面IC考虑了因子值在每个组内的分布情况，从而能够更准确地评估因子与收益率之间的相关性。

使用横截面IC修正后，对于相同IC值的因子A和B，在多头组的因子值与收益率相关性较低的情况下，B因子的横截面IC值会更低，从而更能反映其与收益率之间的实际相关性差异。因此，在进行选股时，应该综合考虑横截面IC值和其它因素，而不是仅仅依赖于单一的IC指标。

来源：ChatGPT，国金证券研究所

图表14: ChatGPT 思维链提示示例 4

WX 若十个组的权重都相等的话，使用横截面IC仍然无法改进。由于我们更看重多头组的因子值与收益率相关性，是否有其他方式改进IC指标

WX 如果所有组的权重相等，那么横截面IC确实无法改进，因为它只考虑了不同组内因子值与收益率的相关性，而没有对不同组的重要性进行加权考虑。

在这种情况下，可以考虑使用加权平均IC (Weighted Average IC) 来改进IC指标。加权平均IC将不同组的IC值加权平均，以反映不同组对整体IC值的贡献程度。通常可以使用某个评价因子（如因子在过去一段时间内的平均收益率或方差等）来作为权重，或者通过根据不同组内股票数量或市值大小来进行加权。

对于这种情况，如果多头组的因子值与收益率相关性较高，那么其加权平均IC值将会更高，从而更能反映其对投资组合收益率的贡献。因此，使用加权平均IC可以更好地评估不同组内因子值与收益率之间的相关性，从而更准确地评估选股因子的有效性。

图表15: ChatGPT 思维链提示示例 5

WX 考虑到A股不能做空的交易限制，应该如何正确给不同组的IC值进行加权

WX 考虑到A股不能做空的交易限制，给不同组的IC值进行加权时需要考虑以下两个方面：

1. 做多组的权重：由于不能做空，只能通过买入股票进行投资，因此在给不同组的IC值进行加权时，应该更加关注做多组的股票。可以考虑使用股票市值或交易量来衡量不同组内做多股票的重要性，然后以加权平均的方式计算做多组的IC值。
2. 做空组的处理：对于不能做空的交易限制，可以将空头组的股票替换为另外一个指数或股票池中的股票。可以选择与空头组中股票风格、行业、市值等相似的股票进行替换，以保持整体投资策略的一致性。

需要注意的是，由于不能做空的交易限制会对投资组合的构建和风险控制产生一定的影响，因此在使用加权平均IC值进行选股时，应该结合实际情况进行综合分析和决策。

来源：ChatGPT，国金证券研究所

来源：ChatGPT，国金证券研究所

可以看出，经过我们的初步提示，ChatGPT 已经认识到我们关注的 IC 指标存在的问题并及时给出了一个解决方案：每组内分别计算 IC 值，再将所有组内的相关系数求均值得到横截面 IC 值。但其逻辑思维能力仍有一定漏洞，我们直接指出后，成功得到了想要的解决方式：计算加权平均 IC 来改进 IC 指标，以反映不同组对整体 IC 值的贡献程度。但由于其对于股票市场 A 股的做空限制可能认识不足，我们进一步提示后，ChatGPT 最终给出了应当给做多组赋予更高权重的回答。但需要注意的是，ChatGPT 被训练时接受了海量的文本和数据，其处理实际问题时仍存在一些“生搬硬套”的现象需要不断提示纠正。

三、ChatGPT 因子挖掘实战

由以上案例可以说明，ChatGPT 在量化研究领域展现出了一定的实用价值。我们此处以常见的因子挖掘作为研究方向，考察经过一定程度的提示 (prompt)，模型是否能给出符合需求的结果。

3.1 周频变异系数因子构建与测试结果

首先我们在中低频上维度上对 ChatGPT 的因子挖掘能力进行测试。此处考虑基于量价数据让 ChatGPT 模型给出一些独特、原创的因子。

图表16: ChatGPT 生成因子示例 4

WX 作为一名资深量化研究员，按照以下指令，构建一个选股因子：

1. 不要使用现有的技术指标去直接构建
2. 构建一个新的、独特的因子
3. 因子的构建逻辑不要太简单
4. 考虑使用价量类数据去研究流动性
5. 请给出构建的详细步骤和计算方法

WX 作为一名资深量化研究员，我将按照以下指令构建一个新的、独特的选股因子，并使用价量类数据去研究流动性。以下是我的详细步骤和计算方法：

1. 我将选择使用股票的价格和成交量数据，因为这些是股票市场中最基本、最重要的数据之一。价格和成交量通常可以反映股票的流动性和市场情绪，也是很多交易策略的基础。
2. 我将计算每个交易日的股票价格的滚动平均值和滚动标准差，并分别表示为MA和STD。同时，我也将计算每个交易日的成交量的滚动平均值和滚动标准差，并分别表示为VMA和VSTD。
3. 接下来，我将计算股票价格的流动性指标LI (Liquidity Index)，它是MA和STD之间的比率，用于衡量价格的波动性和市场流动性。LI的计算公式如下：
 $LI = STD / MA$
4. 我将计算成交量的流动性指标VLI (Volume Liquidity Index)，它是VMA和VSTD之间的比率，用于衡量成交量的波动性和市场流动性。VLI的计算公式如下：
 $VLI = VSTD / VMA$
5. 我将使用LI和VLI两个流动性指标来构建新的选股因子，称为“流动性因子 (LF)”。LF的计算公式如下：
 $LF = LI * VLI$
6. 最后，我将对所有股票进行排名，选取LF值最高的股票作为买入信号，选取LF值最低的股票作为卖出信号。

总体来说，LF因子通过考虑股票价格和成交量的流动性指标，综合反映了股票市场的情况，可以作为一种有效的选股因子，用于寻找市场流动性较好的股票。

来源：ChatGPT，国金证券研究所

该因子在收盘价和成交量两个领域分别计算其变异系数，并相乘得到一个综合的选股因子。用来衡量股票的波动性和

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：国金证券-Alpha掘金系列之五：如何利用ChatGPT挖掘高频选股因子?.pdf

请登录 <https://shgis.com/post/1782.html> 下载完整文档。

手机端请扫码查看：

