

GPT 因子工厂：多智能体与因子挖掘

华泰研究

2024 年 2 月 20 日 | 中国内地

深度研究

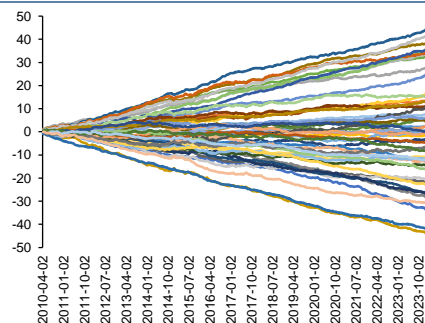
研究员 林晓明
SAC No. S0570516010001 linxiaoming@htsc.com
SFC No. BPY421 +(86) 755 8208 0134

研究员 何康, PhD
SAC No. S0570520080004 hekang@htsc.com
SFC No. BRB318 +(86) 21 2897 2039

人工智能系列之 74：基于 GPT 的多智能体系统应用于量价因子挖掘

本研究介绍大语言模型与智能体相关概念，基于 GPT 和多智能体系统构建端到端的量价因子挖掘系统。在本研究构建的“GPT 因子工厂”中，三类角色各司其职：(1) FactorGPT，负责构建因子表达式；(2) CodeGPT，负责将因子表达式转换为可自动执行的程序，并计算结果；(3) EvalGPT，负责对因子计算结果进行回测检验，同时生成优化建议，并可将建议返回给 FactorGPT，进而继续优化该因子。在二次优化后的因子中，IC 表现靠前（前 20%）绝对值均值为 0.034，RankIC 表现靠前（前 20%）绝对值均值为 0.054，因子表现整体较好。

第二次优化后的累积周度 RankIC



资料来源：Wind, OpenAI, 华泰研究

大语言模型与因子挖掘：大语言模型乘风而上，因子挖掘或可借力腾升

大语言模型指基于大量文本数据预训练的超大型语言模型，其参数规模通常达到数十亿乃至数千亿级别，主流大语言模型大多基于 Transformers 架构构建，通过自监督的预训练方式，获得理解和生成自然语言的能力。大语言模型在诸如办公、绘画、影视、游戏等多个领域的应用持续成熟。Alpha-GPT 是大语言模型应用于量化研究的范例，其通过大语言模型与人类交互的方式赋能因子挖掘，然而 Alpha-GPT 挖掘因子的方式本质依然是遗传算法，没有充分利用大模型处理数据、生成代码、调用工具等能力，而这些能力恰好匹配因子挖掘各项流程的关键点。

大语言模型多智能体系统：多智能体分工合作展现群体智慧

智能体概念最早来源于亚里士多德等人的哲学思考。通常而言，智能体指具有行动能力的实体，具有行使意志、做出选择、采取行动、对外界刺激做出反应的能力。大语言模型具备智能体的反应性、主动性和社交能力等特征，总体符合智能体的定义，具备成为通用人工智能的潜力。更进一步，让不同的智能体负责不同的分工，也即多智能体系统，通过协商合作的方式或能更加有效地解决复杂问题。相较于单智能体，多智能体系统具有更加广阔的应用前景，也在大语言模型研究领域受到了更多的关注。

因子挖掘效果：GPT 因子工厂展现出高水平的因子挖掘效果

本文对“GPT 因子工厂”的因子挖掘效果进行测试，进行 50 次因子挖掘，每次因子挖掘包括首次挖掘、第一次优化和第二次优化。首次因子挖掘获得的因子中，分层 1 年化超额收益均值为 11.14%，分层 5 年化超额收益均值为 -1.11%，分层 1-分层 5 年化超额收益均值为 12.25%，因子分层效果总体较好。因子相关性方面，首次因子挖掘中，因子相关系数绝对值均值为 0.229，优化后变为 0.192 和 0.230，第一次优化后与第二次优化后的因子相关性水平依旧保持，因子工厂产出的因子相关性普遍较低。

因子优化效果：优化后的因子效果提升明显

EvalGPT 可根据回测结果对单个因子提出优化建议，并提交建议至 FactorGPT 用于优化因子表达式。从结果上看，50 次测试中，第二次优化后，相较首次挖掘和第一次优化，显示出较为明显的提升效果，例如 $|t|$ 均值从 4.57 和 4.53 提升至 4.65， $|t| > 2$ 占比从 69.44% 和 69.60% 提升至 71.27%。优化建议内容方面，词频结果显示 EvalGPT 擅长从窗长等角度对因子表达式提出优化建议。

风险提示：GPT 挖掘因子是对历史的总结，具有失效风险。GPT 挖掘量价因子可解释性受限，使用需谨慎。大模型训练集广泛，可能存在过拟合风险。

正文目录

引言3
 大语言模型与多智能体4
 大语言模型与因子挖掘5
 基于大语言模型的智能体5
 反应性6
 主动性6
 社交能力6
 从单智能体到多智能体6
 LangChain：多智能体开源实现框架6
 方法8
 GPT 因子工厂8
 FactorGPT8
 CodeGPT8
 EvalGPT9
 测试方案9
 结果10
 因子挖掘实例10
 因子挖掘效果11
 因子相关性12
 因子优化效果14
 因子释义与优化建议15
 总结16
 参考文献17
 风险提示17

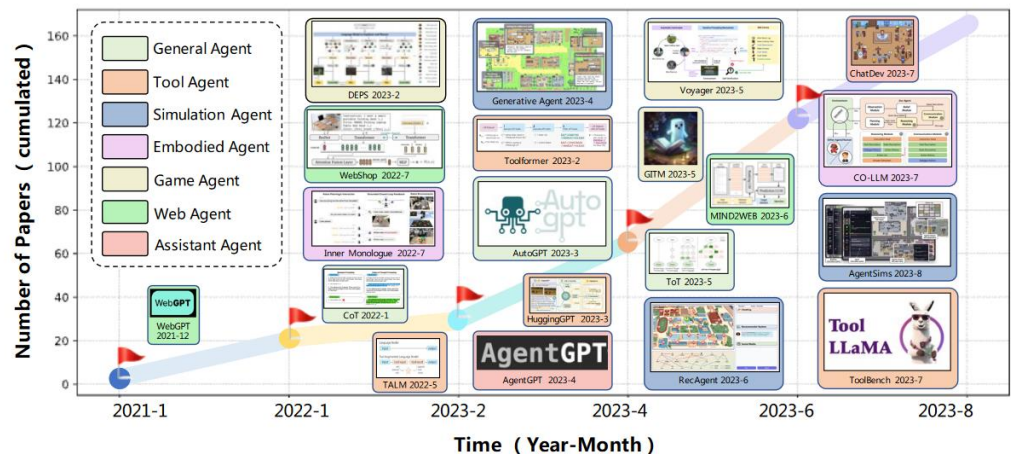
导言

“积力之所举，则无不胜也；众智之所为，则无不成也。”——《淮南子·主术训》

自 OpenAI 推出 ChatGPT 以来，大语言模型（Large Language Model, LLM）在语义理解、多轮对话、内容生成等方面展现出的能力频频超出预期，通用人工智能（Artificial General Intelligence, AGI）似乎已处目之所及之处。大语言模型在诸如办公、绘画、影视、游戏等多个领域的应用持续成熟，而量化金融领域的大语言模型应用却寥寥无几。如何以大语言模型为底座，实现有效且完善的量化研究应用，可能是值得探索的通幽之径。

在大语言模型领域中，模型本身的能力尚显单薄，为了进一步发挥大语言模型的能力，使其能自动化地完成高复杂度任务，学界持续探索基于大模型的多智能体（Multi-Agent）框架。多智能体系统由多个智能体组成，每个智能体的实现均依赖于单个大语言模型，并具有与环境交互、与其他智能体合作协同的能力。智能体可以通过一系列的连续行为，实现人类给定的任务目标。有学者认为，基于大语言模型的多智能体可能是迈向通用人工智能重要途径之一。

图表1：基于大语言模型的自主性智能体发展趋势



资料来源：Wang et al. (2023) A Survey on Large Language Model based Autonomous Agents，华泰研究

在量化研究中，与多智能体系统较为贴近的应用场景之一可能是因子挖掘。传统的因子挖掘框架往往基于遗传规划，涉及算子组合、回测优化等多项子流程，整体过程较为复杂，直接调用大语言模型难以完成；但基于多智能体系统，使大模型调用工具、与环境交互、以及智能体间分工合作，或能显著提升大语言模型完成复杂任务的能力。

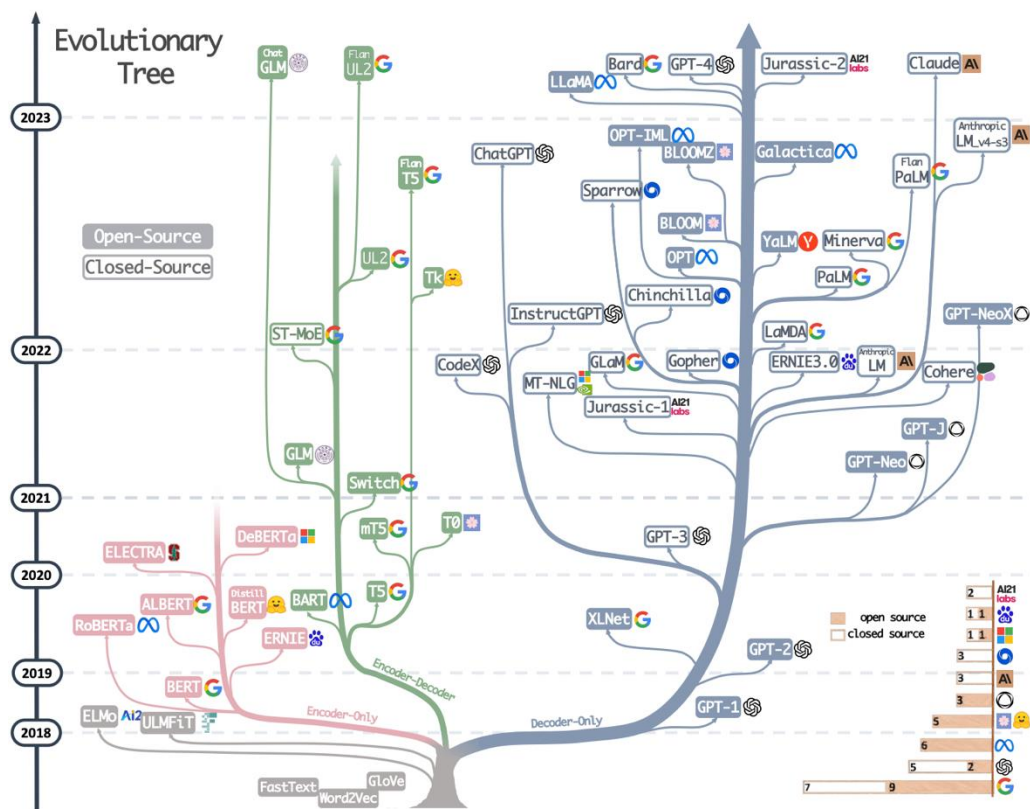
本研究基于 GPT 模型，通过大模型扮演“GPT 因子工厂”中的三个不同角色，以实现因子挖掘功能：FactorGPT 负责构建和优化因子表达式，CodeGPT 负责生成代码文件，EvalGPT 负责因子回测和评估，最终实现端到端因子挖掘系统。我们使用“GPT 因子工厂”进行 50 次因子挖掘测试，每次因子挖掘包括首次挖掘、第一次优化和第二次优化。测试结果显示：因子表现方面，首次挖掘中，IC 表现靠前（前 20%）绝对值均值为 0.031，RankIC 表现靠前（前 20%）绝对值均值为 0.045，首次挖掘出的因子表现整体较好；因子相关性方面，首次因子挖掘中，因子相关系数绝对值均值为 0.229，第一次优化后变为 0.192，第二次优化后变为 0.230，总体相关性偏低；因子优化方面，第二次优化后的因子相较首次挖掘和第一次优化，显示出较为明显的提升效果。

大语言模型与多智能体

大语言模型（Large Language Model, LLM）指基于大量文本数据预训练的超大型语言模型，其参数规模通常达到数十亿乃至数千亿级别。目前主流大语言模型都基于 Transformers 架构，通过自监督的预训练方式，获得理解和生成自然语言的能力。经典的 Transformers 结构分为编码器（Encoder）和解码器（Decoder）两部分，大语言模型由此可以分为三种架构：第一种是仅使用编码器（Encoder-Only），代表模型如 BERT；第二种是仅使用解码器（Decoder-Only），代表模型如 GPT；第三种是同时使用编码器和解码器（Encoder-Decoder），代表模型如 T5。

随着模型规模的增长，使用解码器的模型展现出更强大的文本生成能力，逐渐受到研究者的青睐，如图表 2 所示。OpenAI 在 2020 年发布了具有 1750 亿参数的 GPT-3，其强大的多任务和情景学习（In-Context Learning）的能力标志着自然语言处理进入新的研究阶段。在此基础上，人类采取指令微调（Instruction Fine-Tuning）、基于人类反馈的强化学习（Reinforcement Learning from Human Feedback）等方法，让模型能够遵循指令、遵守道德，显著提高了模型理解、推理、生成等能力。OpenAI 在 2022 年发布的 ChatGPT 可以通过自然语言和人类交互、完成指定任务，包括撰写文稿、编写代码、提供建议等，引起大语言模型新的发展热潮。

图表2：大语言模型进化树



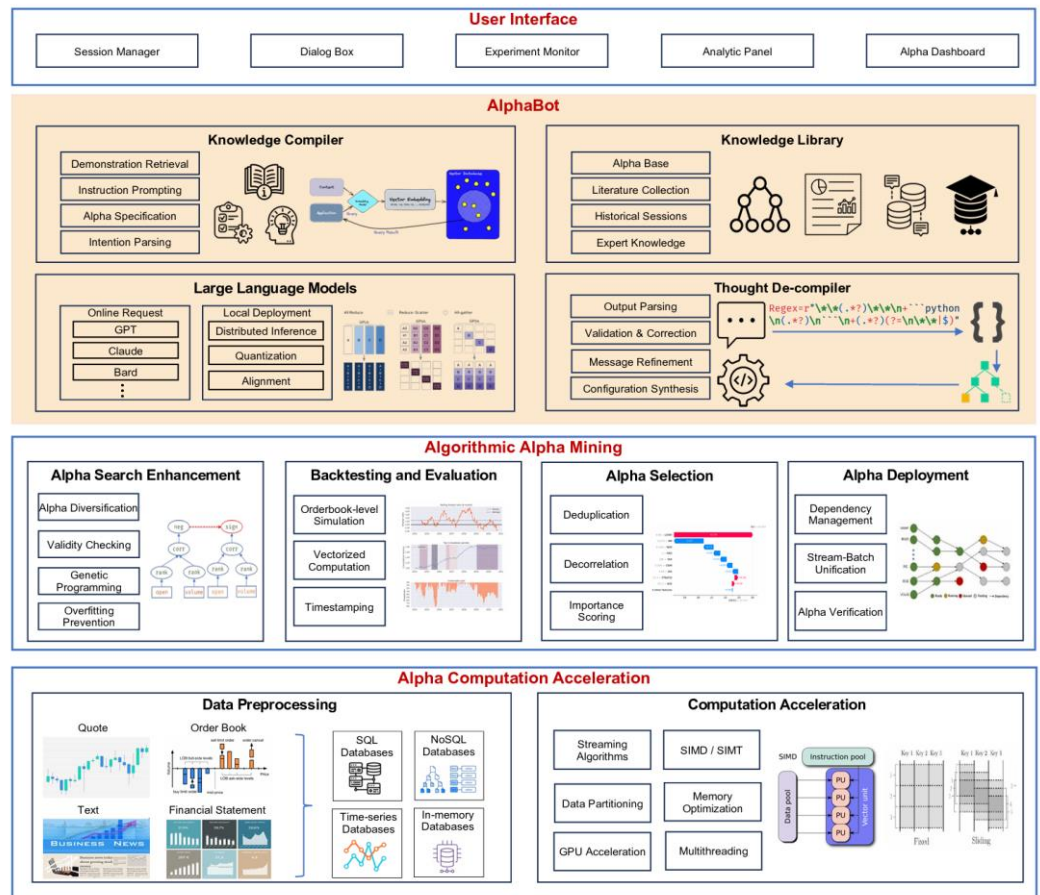
资料来源：Yang et al. (2023) Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond，华泰研究

大语言模型虽然能力出众，但其训练和运行成本较为高昂，普通用户很难私有化部署；此外，目前开源的大语言模型在性能上与闭源模型依然有显著差距。因此大模型公司通常采用“语言模型即服务”（LLM-as-a-Service）的方式，通过网页或者接口为用户提供服务，例如百度于 2023 年推出文心一言大语言模型，允许用户通过网页和接口访问该模型。

大语言模型与因子挖掘

目前大语言模型在多个垂直领域已经开始落地应用，例如金融领域大模型 BloombergGPT、医疗领域大模型灵医大模型、政务领域九州大模型等。在量化研究中，Saizhuo Wang 等人提出了 Alpha-GPT，通过大语言模型与人类交互的方式赋能因子挖掘。他们将大语言模型作为量化研究员和因子挖掘之间的中间层：利用大模型将研究员的意图转化为种子因子，再通过传统的遗传算法搜索有效因子，通过回测和评估的因子会通过大模型进行解释，最终呈现给研究员。在整个过程当中，大模型的作用主要是辅助人工，简化了研究员的工作流程。

图表3：Alpha-GPT 因子挖掘框架



资料来源：Wang et al. (2023) Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment, 华泰研究

然而，Alpha-GPT 挖掘因子的方式本质依然是遗传算法，没有充分利用大模型处理数据、生成代码、调用工具等能力。与之不同，本文期望充分利用大模型的上述能力，基于大模型实现端到端的量价因子挖掘框架，实现因子从构思到回测再到优化的自动化全流程。

基于大语言模型的智能体

智能体的概念由来已久，最早来源于亚里士多德等人的哲学思考。通常而言，智能体是指具有行动能力的实体，具有行使意志、做出选择、采取行动、对外界刺激做出反应的能力。有研究者认为，大语言模型已经具有智能体的一些特征，包括反应性、主动性和社交能力。

反应性

反应性是指智能体可以对环境中的刺激和变化作出迅速的反应。大语言模型可以感知文本输入的变化，并通过文本输出作出相应的反馈。此外，人们还通过多模态技术扩展了大语言模型的感知和行为空间，例如，让模型具备调用工具的能力，从而与现实世界进行交互；具备调用工具能力的大语言模型可以通过搜索引擎检索金融知识；通过计算器计算复杂的公式；从数据库中读取数据等等。

OpenAI 提供的多个 GPT 模型都具有调用工具的能力。使用者需要在 API 接口中提供作为工具的函数的功能描述和接口信息，当模型认为需要调用函数时，会返回特定的字段，其中包括调用函数的具体参数；在函数执行结束之后，用户还可以将调用结果以文本的形式反馈给模型，供其选择下一步行动。

主动性

主动性是指智能体具备主动采取目标导向的行动以实现目标的能力。智能体不仅可以对环境的刺激做出反馈，还可以推理、规划、采取积极行动以达到指定目标。大语言模型已经展现出很强的推理和规划能力，包括逻辑推理能力、数学推导能力、任务分解能力等。

社交能力

社交能力是指智能体可以与其他智能体（包括人类）通过某些语言进行交互。大语言模型已经展现出较强的自然语言理解和生成能力，可以轻松地与人类进行沟通；此外，人们发现利用一些特殊的提示，可以使大语言模型扮演不同的角色，模仿现实社会中的劳动分工。通过让多个智能体模仿社会中的合作与竞争，人们可以提高大语言模型完成复杂任务的能力。

从单智能体到多智能体

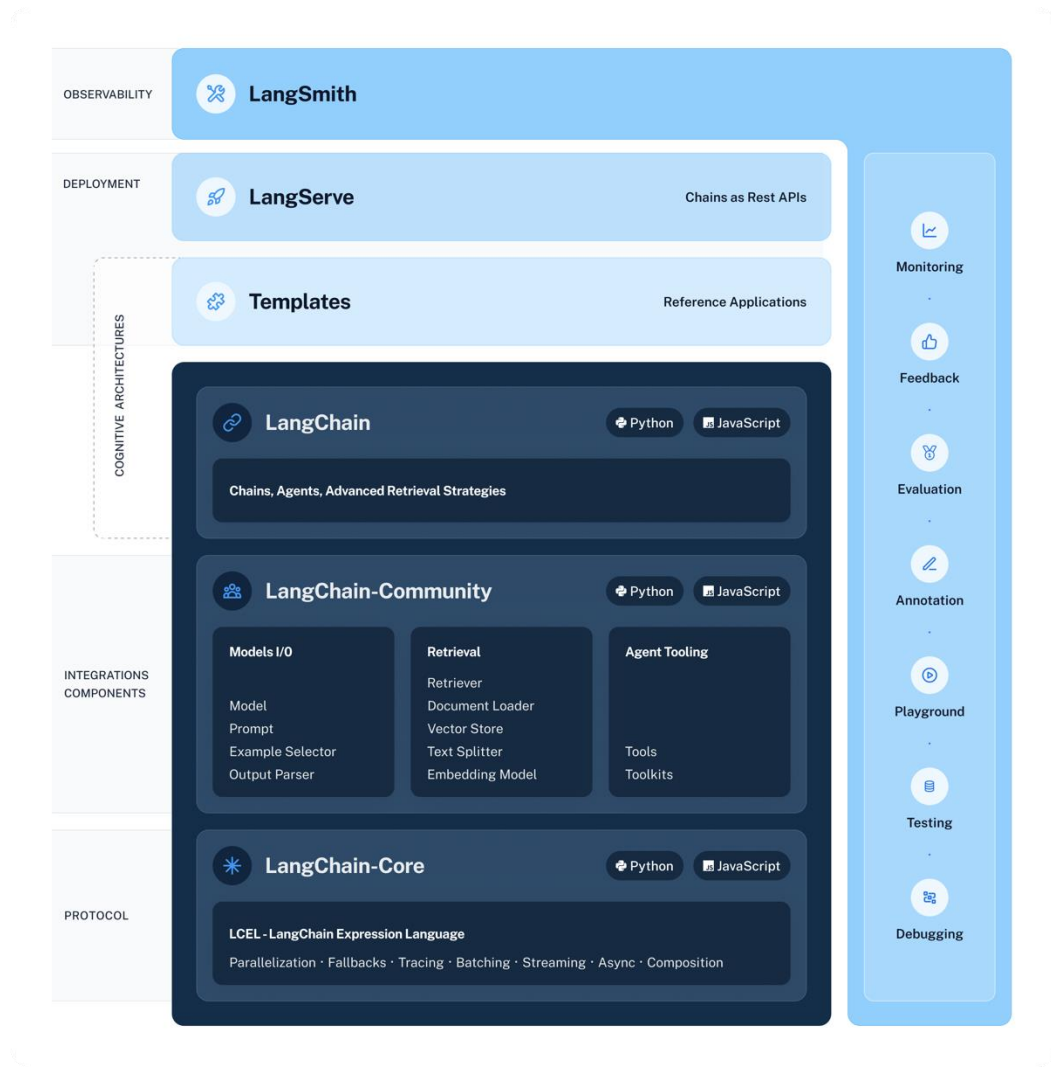
亚当·斯密在《国富论》中曾说：“劳动生产力上最大的增进，以及运用劳动时所表现更大的熟练、技巧和判断力，似乎都是分工的结果”。基于大模型的单智能体虽然可以出色地完成理解、生成任务，但是不能与其他智能体进行沟通合作。而基于大模型的多智能体系统则可以体现出“群体的智慧”：让不同的智能体负责不同的分工，通过协商合作的方式更加有效地解决问题。一方面，分工可以让每个智能体专注于解决特定问题，更好地释放其完成任务的潜力；另一方面，智能体之间多轮的沟通可以让它们在解决问题时更加全面深入，提高了他们解决问题的能力。因此，相较于单智能体，多智能体系统具有更加广阔的应用前景，也受到了更多的研究关注。

总的来说，大语言模型的多项能力符合智能体的定义，具备成为通用人工智能的潜力。利用大语言模型搭建智能体，充分发挥大语言模型完成复杂任务的能力，成为目前人工智能领域的研究热点之一。

LangChain：多智能体开源实现框架

LangChain 是一个开源的 Python 框架，旨在利用大语言模型开发端到端的应用程序。通过 LangChain 提供的模块、工具，开发者可以将大语言模型集成到应用程序之中，大大降低了基于大语言模型的应用开发门槛。

图表4：LangChain 架构



资料来源：LangChain，华泰研究

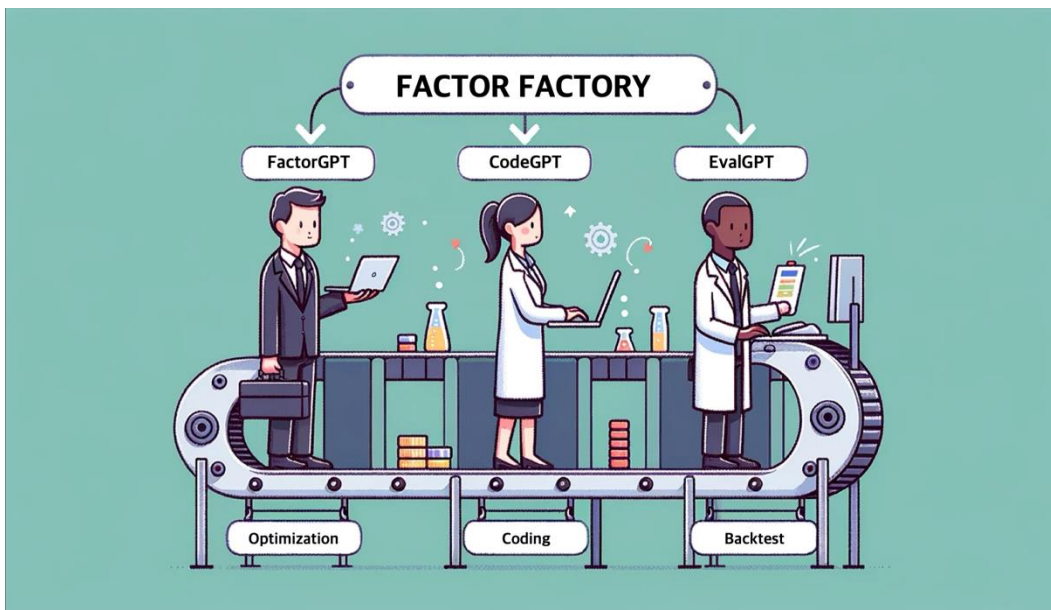
LangChain 框架覆盖了大语言模型应用从开发到产品化再到部署的全过程，目前分为 4 个部分：LangChain Libraries 是 Python 和 Javascript 库，包含各模块的接口和集成；LangChain Templates 包含了大量任务的提示模版；LangServe 可以直接将 LangChain 中的模块部署为 REST 接口；LangSmith 为开发者提供了调试、测试、评估平台。LangChain 采用模块化设计，可以组合不同的模块实现自定义的功能。我们可以自定义提示模版，将动态信息转换为适合大语言模型的格式；可以设计输出解析器，将大语言模型的输出转换为格式化信息；可以定义智能体和工具组，根据用户的输入决定调用合适的工具。

LangChain 是目前影响力最大的大语言模型应用开发框架之一，本研究参考了 LangChain 的部分设计思路，构建起基于多智能体的因子挖掘框架。

方法

本研究设计构建的“GPT 因子工厂”是基于 GPT 模型的多智能体系统，在该系统中，智能体具备三种不同分工：（1）FactorGPT，负责构建因子表达式；（2）CodeGPT，负责将因子表达式转换为可自动执行的程序，并计算结果；（3）EvalGPT，负责对因子计算结果进行回测检验，同时生成优化建议，并将建议返回给 FactorGPT，进而继续优化该因子。因子挖掘过程中，所有步骤均可由智能体自动执行，无需人为干预，最终将输出因子的表达式、含义、代码、回测结果以及优化建议。

图表5：GPT 因子工厂示意图



资料来源：OpenAI，华泰研究

GPT 因子工厂

FactorGPT

FactorGPT 在“因子工厂”中负责利用底层数据字段生成因子表达式。为提高模型的生成质量，我们采用了 Few-Shot 方式，在提示中加入已有的因子示例。我们要求模型直接输出因子的名称、数学表达式和含义。为了获得稳定的输出格式，我们使用输出解析器将输出的因子信息解析为结构化内容。

图表6：FactorGPT 使用字段

字段	含义
S_DQ_OPEN	开盘价(元)
S_DQ_HIGH	最高价(元)
S_DQ_LOW	最低价(元)
S_DQ_CLOSE	收盘价(元)
S_DQ_VOLUME	成交量(手)

资料来源：Wind，华泰研究

CodeGPT

CodeGPT 负责将因子表达式转换为可执行的代码文件。为降低模型“幻觉”现象，我们设定模型可自主调用或自行生成算子，使得代码可执行性更高。模型生成代码后，将运行代码文件，若有报错信息抛出，模型将根据报错信息修改代码文件。模型将不断迭代上述过程直到代码文件可以运行，直至成功运行。

EvalGPT

EvalGPT 负责对因子的计算结果进行回测，并根据回测结果给出评价和优化建议。EvalGPT 的输入为因子值，输入之后将通过回测模块获得因子回测结果，例如 IC、RankIC、年化收益率等指标。随后，EvalGPT 会根据回测结果对因子做出评价和优化建议。为了获得稳定的输出格式，我们同样使用输出解析器进行内容解析。

在经历了上述三个智能体处理后，我们将获得如下信息：因子名称、因子表达式、因子含义、因子可执行代码文件、因子值、因子回测结果、因子评估结果、因子优化建议。如需继续优化该因子，我们可将 EvalGPT 生成的优化建议直接反馈给 FactorGPT，让大模型继续改进当前因子，并重复上述流程。通过以上设计，我们实现了因子“构建-计算-回测-优化”的全自动循环流程。

测试方案

为检验“GPT 因子工厂”因子挖掘的能力，本研究进行 50 次因子挖掘测试。每次因子挖掘包含三次全流程循环：

- （1）首次因子挖掘：FactorGPT 生成因子表达式，CodeGPT 生成代码，EvalGPT 进行回测并提出优化建议；
- （2）第一次优化：FactorGPT 根据上一轮 EvalGPT 提出的优化建议改进因子表达式，CodeGPT 和 EvalGPT 继续重复其分工流程；
- （3）第二次优化：FactorGPT 再次根据 EvalGPT 提出的建议改进表达式，CodeGPT 和 EvalGPT 继续重复工作。

结果

因子挖掘实例

以下展示 GPT 因子工厂的运行实例。如图表 7，GPT 因子工厂运行过程中，将持续汇报当前因子挖掘进展，便于详细观察多智能体 FactorGPT、CodeGPT 和 EvalGPT 的工作内容。

图表7：GPT 因子工厂运行实例截图

```
【GPT因子工厂】第1次挖掘开始
FactorGPT:
【因子名称】
量价动量因子(Price-Volume Momentum Factor)

【因子计算方法】
Rank_Mul(Decay_Linear(TS_Return(S_DQ_CLOSE, 10), 10), Rank(Div(S_DQ_VOLUME, TS_Mean(S_DQ_VOLUME, 10))))

【因子含义】
该因子通过结合收盘价变化率的加权平均(即短期动量)和成交量变化相对于其短期均值的排名(即成交量动量),构建了一个量价动量因子。具体地,首先计算过去10个交易日收盘价的变化率,并对这些变化率
CodeGPT:
正在调用函数: get_calculator_function
算子名称: Rank_Mul
CodeGPT:
正在调用函数: get_calculator_function
算子名称: Decay_Linear
CodeGPT:
正在调用函数: get_calculator_function
算子名称: TS_Return
CodeGPT:
正在调用函数: get_calculator_function
算子名称: Rank
```

资料来源：OpenAI，华泰研究

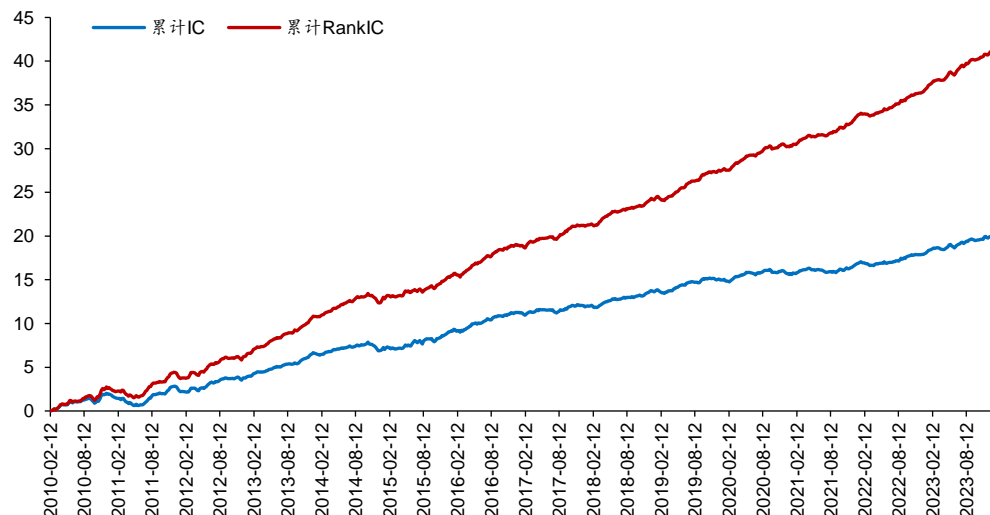
图 8、图 9 和图 10 展示 GPT 因子工厂挖掘出因子后，进行第二次优化后的一个因子实例。从累计周度 IC 和 RankIC 看，因子表现较为稳定，方向为正向；从分层回测结果上看，因子分层差异明显，不同分层间指标单调变化，表明该因子效果较好。

图表8：GPT 因子工厂因子实例

因子名称 (FactorGPT)	因子释义 (FactorGPT)	因子表达式 (FactorGPT)
	该因子进一步优化了量价关系并强调了价格波动的倒数,以突出平均交易量大但价格波动性较低的股票。首先计算过去 10 天股价的移动平均值,再计算过去 10 天成交量的移动平均值,然后将这两个移动平均值进行排名处理,并计算它们的排名比值。排名比值旨在识别那些股价持续稳定上涨且成交量较高的股票。同时考虑了过去 30 天股票最高价与最低价之差的标准差的逆,从而减小了高波动股票在因子中的权重。	Rank_Mul(Div(Rank(TS_Mean(S_DQ_CLOSE, 10)), Rank(TS_Mean(S_DQ_VOLUME, 10))), Inv(TS_Std(Sub(S_DQ_HIGH, S_DQ_LOW), 30)))
更进一步优化的量价动力因子	通过这种设计,因子着重关注那些具有稳定上涨动力和相对较高的市场活跃度,但又不受日内过度波动影响的股票。这不仅有助于提升因子的预测准确性,还有助于筛选出那些潜在的低风险且收益稳定的投资机会。	
	这样的进一步改进可以: 1. 通过稳定的移动平均来降低短期波动的影响,提升因子的稳定性。 2. 通过捕获长期的量价关系来增强因子的预测能力。 3. 引入价格波动的倒数来降低高波动性股票的影响,寻找波动性小、趋势稳定的交易机会。 4. 改善 IC 标准差和 rank_IC 标准差,提升整体因子的信息比率。	

资料来源：Wind，OpenAI，华泰研究

图表9：GPT 因子工厂因子累计周度 IC 和 RankIC 实例



注：GPT 命名该因子为“更进一步优化的量价动力因子”，回测区间：2010-01-04 至 2024-02-08

资料来源：Wind, OpenAI, 华泰研究

图表10：GPT 因子工厂因子分层回测结果实例

	年化收益率	年化波动率	夏普比率	最大回撤	年化超额收益 率	超额收益年化 波动率	信息比率	超额收益最大 回撤	调仓胜率	相对基准盈亏 比
分层 1	20.20%	23.11%	0.87	-39.97%	18.60%	11.00%	1.69	-25.40%	62.66%	1.03
分层 2	13.83%	26.43%	0.52	-52.85%	13.30%	10.57%	1.26	-23.51%	60.42%	1.01
分层 3	7.89%	27.18%	0.29	-60.47%	7.71%	9.70%	0.79	-22.55%	58.88%	0.94
分层 4	0.05%	28.33%	0.00	-70.76%	0.17%	9.99%	0.02	-42.23%	54.55%	0.86
分层 5	-8.37%	27.92%	-0.30	-86.21%	-8.42%	10.49%	-0.80	-71.25%	46.57%	0.88
基准	0.87%	22.57%	0.04	-57.85%						

注：GPT 命名该因子为“更进一步优化的量价动力因子”，回测区间：2010-01-04 至 2024-02-08

资料来源：Wind, OpenAI, 华泰研究

因子挖掘效果

首次因子挖掘能最大程度反映 GPT 在因子挖掘方面的能力。图表 11 展示初次挖掘出的部分因子表达式，因子表达式总体较为复杂，FactorGPT 可以产出复杂且具备因子释义的因子。对于全部因子回测结果，从累计周度 IC 和累计周度 RankIC 看，GPT 挖掘出的因子方向较为稳定，不乏持续展现出 IC 和 RankIC 持续单调且低波动的因子。

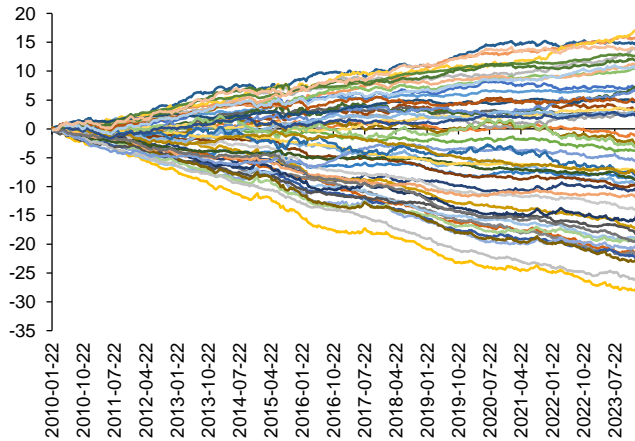
图表11：GPT 因子工厂初次挖掘产出的部分因子表达式

因子表达式 (FactorGPT)

GPT-Alpha1	Div(TS_Corr(Mul(S_DQ_CLOSE, S_DQ_VOLUME), TS_Return(S_DQ_CLOSE, 1), 10), TS_Std(S_DQ_VOLUME, 10))
GPT-Alpha2	Div(TS_Return(Add(S_DQ_HIGH, S_DQ_LOW), 10), TS_Mean(S_DQ_VOLUME, 10))
GPT-Alpha3	Rank_Div(TS_Corr(Decay_Linear(S_DQ_CLOSE, 5), Decay_Linear(S_DQ_VOLUME, 5), 5), TS_Std(Sub(Add(S_DQ_HIGH, S_DQ_LOW), Mul(S_DQ_OPEN, 2)), 5))
GPT-Alpha4	Rank_Mul(TS_Return(S_DQ_CLOSE, 20), Rank(TS_Mean(S_DQ_VOLUME, 20)))
GPT-Alpha5	Rank_Div(Rank_Add(TS_Return(S_DQ_CLOSE, 10), TS_Std(Sub(S_DQ_HIGH, S_DQ_LOW), 10)), Rank(TS_Mean(S_DQ_VOLUME, 10)))
GPT-Alpha6	Rank_Div(Rank_Add(TS_Std(S_DQ_CLOSE, 20), TS_Std(S_DQ_VOLUME, 20)), Rank_Mul(TS_Return(S_DQ_CLOSE, 5), Inv(TS_Corr(S_DQ_CLOSE, S_DQ_VOLUME, 5))))
GPT-Alpha7	Rank_Sub(Rank(Div(Decay_Linear(TS_Std(S_DQ_CLOSE, 20), 20), Decay_Linear(TS_Std(S_DQ_VOLUME, 20), 20))), Rank(TS_Corr(Decay_Linear(S_DQ_CLOSE, 20), Decay_Linear(S_DQ_VOLUME, 20), 20)))
GPT-Alpha8	Rank_Add(Rank_Sub(TS_Std(S_DQ_CLOSE, 20), TS_Std(S_DQ_VOLUME, 20)), Rank_Mul(TS_Return(Add(S_DQ_HIGH, S_DQ_LOW), 5), Inv(TS_Corr(S_DQ_CLOSE, S_DQ_VOLUME, 5))))
GPT-Alpha9	Rank_Add(Rank_Mul(Decay_Linear(TS_Return(S_DQ_CLOSE, 10), 10), Inv(TS_Corr(S_DQ_CLOSE, S_DQ_VOLUME, 10))), Rank_Sub(Rank(Div(TS_Std(S_DQ_HIGH, 20), TS_Std(S_DQ_LOW, 20))), Rank(TS_Std(S_DQ_VOLUME, 20)))
GPT-Alpha10	Rank_Add(Rank_Div(TS_Return(S_DQ_CLOSE, 10), TS_Mean(S_DQ_VOLUME, 10)), Rank(TS_Std(Add(S_DQ_CLOSE, Mul(Delay(S_DQ_CLOSE, 1), -1)), 10)))

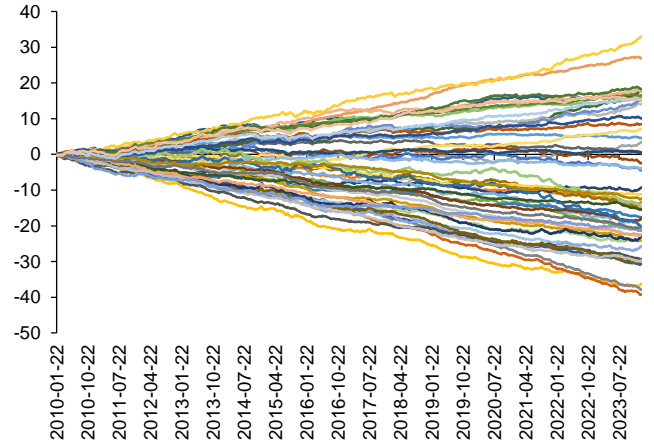
资料来源：Wind, OpenAI, 华泰研究

图表12：首次因子挖掘因子累计周度 IC



资料来源：Wind, OpenAI, 华泰研究

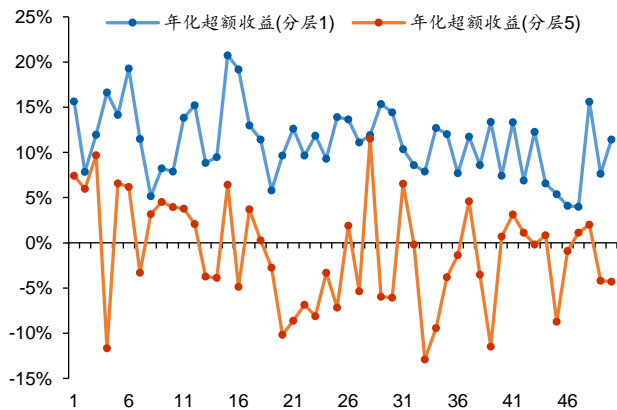
图表13：首次因子挖掘因子累计周度 RankIC



资料来源：Wind, OpenAI, 华泰研究

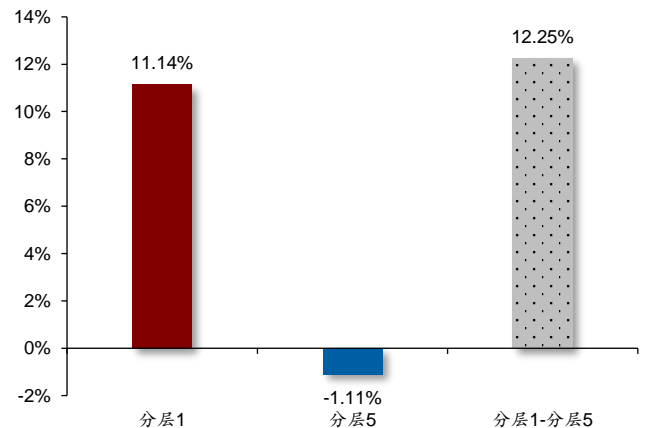
从分层回测结果上看，50次因子挖掘中，因子分层1年化超额收益普遍高于分层5超额收益，分层1-分层5最大值为28.29%，最小值为0.40%。从分层回测年化超额收益均值上看，分层1年化超额收益均值为11.14%，分层5年化超额收益均值为-1.11%，分层1-分层5年化超额收益均值为12.25%。总体来看，因子分层效果较好。

图表14：首次因子挖掘分层回测年化超额收益



注：基准指数为中证全指，因子方向均为调整后
资料来源：Wind, OpenAI, 华泰研究

图表15：首次因子挖掘分层回测年化超额收益均值



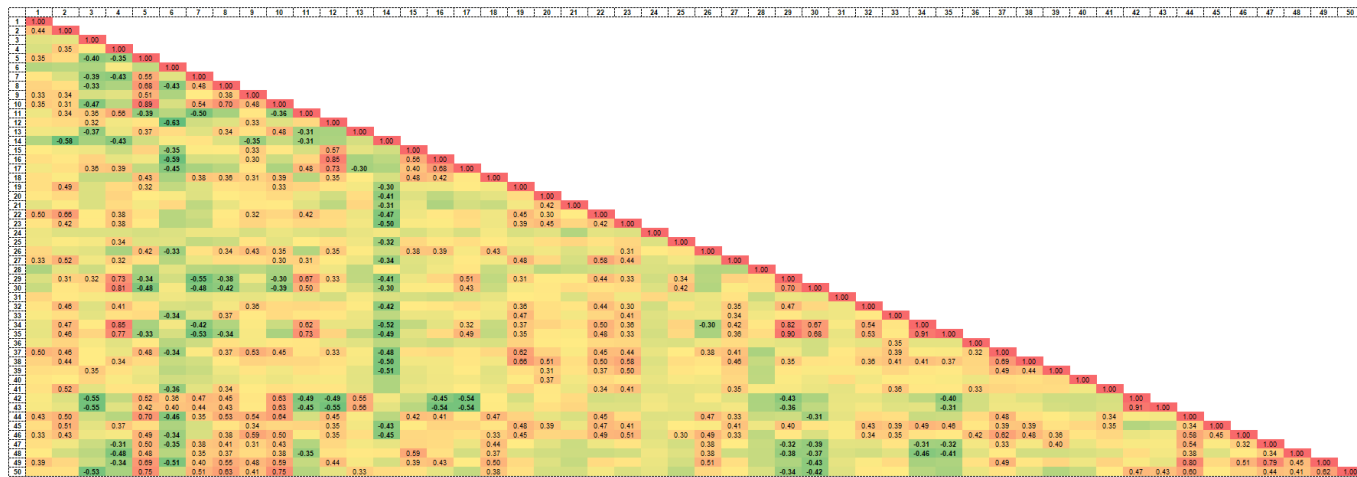
注：基准指数为中证全指，因子方向均为调整后
资料来源：Wind, OpenAI, 华泰研究

因子相关性

因子相关性对于评价因子挖掘框架而言尤为重要。我们统计 GPT 因子工厂进行 50 次因子挖掘的因子相关系数，结果如下图。对于首次因子挖掘，因子相关系数最大值为 0.911，最小值为 -0.632，因子相关为负的相关系数均值为 -0.186，因子相关为正的相关系数均值为 0.248，因子相关系数绝对值均值为 0.229。平均来看，50 次因子挖掘出的因子相关性处于较低水平。

第一次优化后，因子相关系数最大值为 0.753，最小值为 -0.745，因子相关为负的相关系数均值为 -0.175，因子相关为正的相关系数均值为 0.202，因子相关系数绝对值均值为 0.192；第二次优化后，因子相关系数最大值为 0.835，最小值为 -0.872，因子相关为负的相关系数均值为 -0.224，因子相关为正的相关系数均值为 0.236，因子相关系数绝对值均值为 0.230。总体而言，优化后因子相关系数极值在正负区间更为均衡，第一次优化后因子相关性均值水平有所降低，但第二次优化相比于第一次优化并无提升。

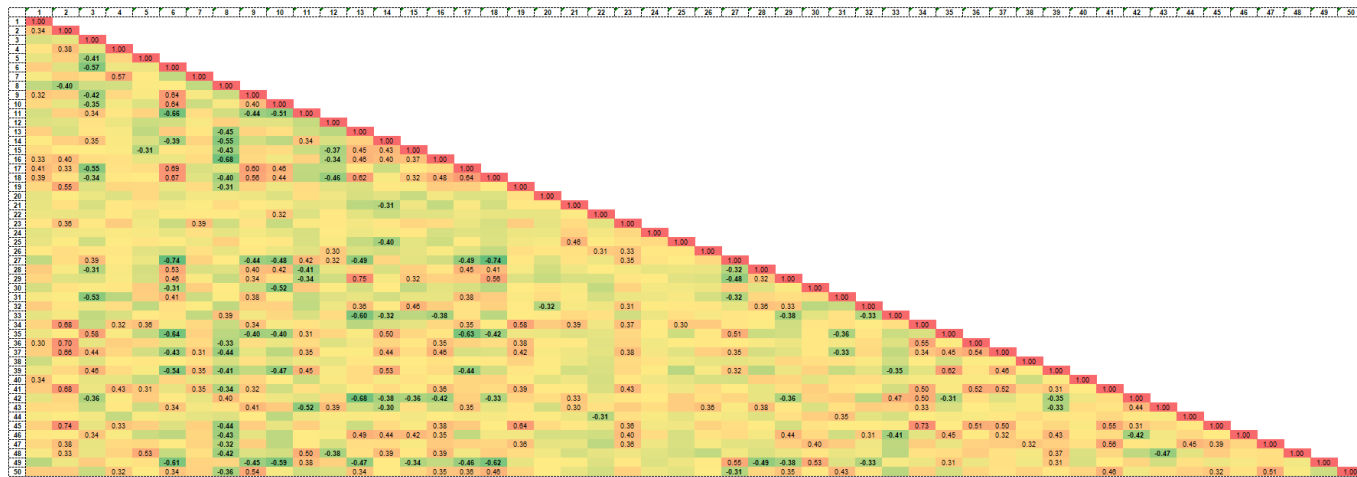
图表16：首次因子挖掘因子相关性



注：相关系数处于-0.3~0.3之外的标注准确数字，否则不标注准确数字

资料来源：Wind, OpenAI, 华泰研究

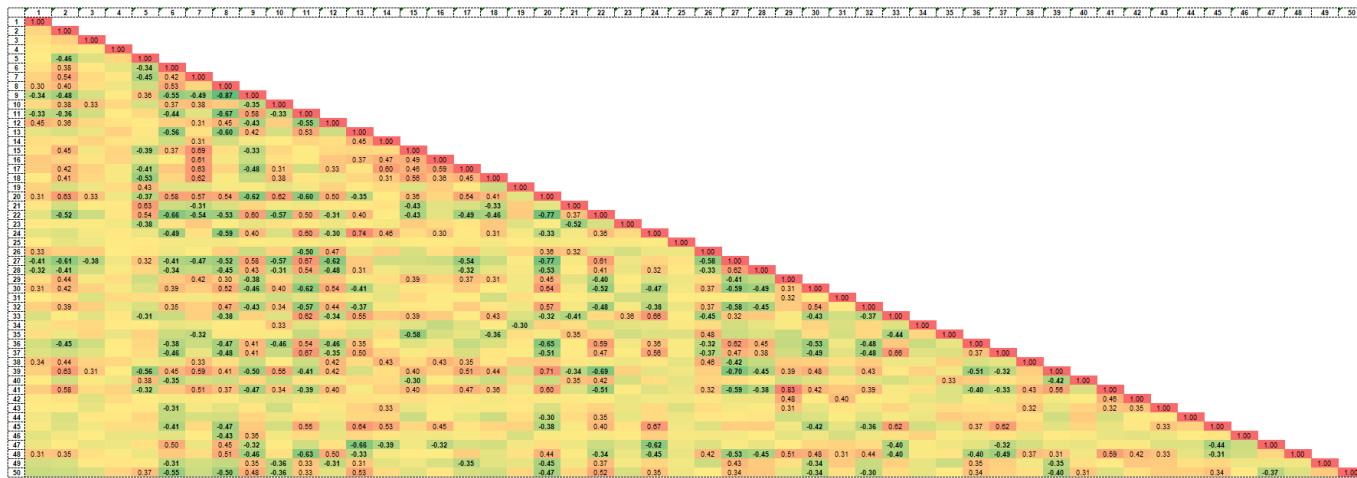
图表17：第一次优化后因子相关性



注：相关系数处于-0.3~0.3之外的标注准确数字，否则不标注准确数字

资料来源：Wind, OpenAI, 华泰研究

图表18：第二次优化后因子相关性



注：相关系数处于-0.3~0.3之外的标注准确数字，否则不标注准确数字

资料来源：Wind, OpenAI, 华泰研究

图表19：因子相关性水平统计汇总

	最大值	最小值	负相关均值	正相关均值	绝对值均值
首次因子挖掘	0.911	-0.632	-0.186	0.248	0.229
第一次优化后	0.753	-0.745	-0.175	0.202	0.192
第二次优化后	0.835	-0.872	-0.224	0.236	0.230

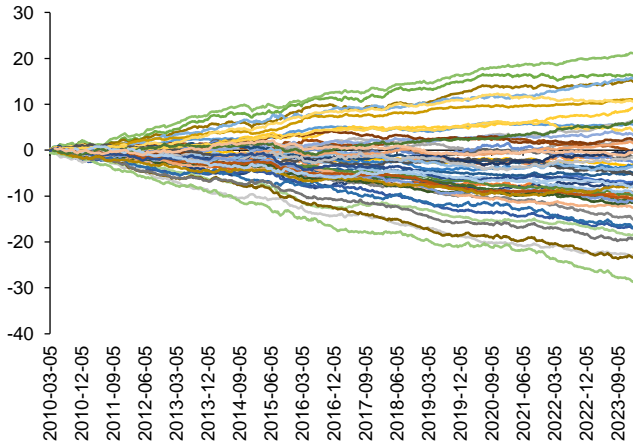
资料来源：Wind, OpenAI, 华泰研究

因子优化效果

GPT 因子工厂的灵活性之一在于能够进行无限次优化，我们在 50 次因子挖掘测试中，按顺序进行两次连续优化，我们预期 EvalGPT 能够根据回测结果进行更具针对性的因子优化。从累计 IC 和 RankIC 曲线分布的形态上看，第一次优化和第二次优化后，50 个因子的曲线分布更加扩散，意味着优化后的因子 IC 和 RankIC 存在一定提升。数值上也进一步确认了这一点：首次挖掘因子累计周度 IC 最大值为 17.23，最小值为-27.67，累计周度 RankIC 最大值为 32.90，最小值为-39.27；第一次优化后因子累计周度 IC 最大值为 21.10，最小值为-28.52，累计周度 RankIC 最大值为 30.38，最小值为-49.75；第二次优化后因子累计周度 IC 最大值为 29.86，最小值为-28.04，累计周度 RankIC 最大值为 44.11，最小值为-43.36。

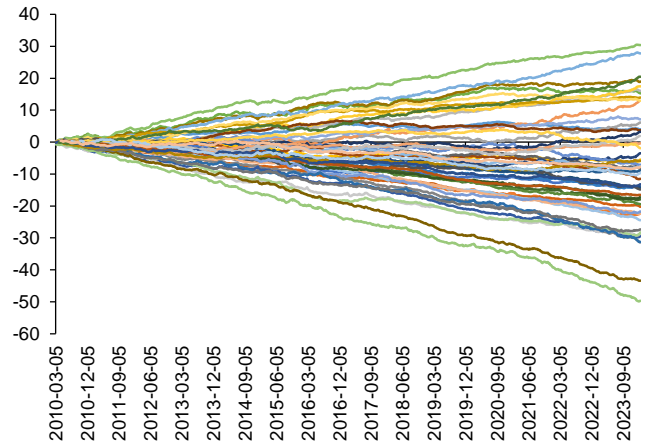
EvalGPT 在第二次优化后的部分相关指标上，相较首次挖掘和第一次优化，显示出较为明显的提升效果。例如第二次优化后，IC 均值绝对值前 20%求平均从 0.031（首次挖掘，下同）和 0.028（第一次优化后，下同）提升至 0.034，RankIC 均值绝对值前 20%求平均从 0.045 和 0.046 提升至 0.054， $|t|$ 均值从 4.57 和 4.53 提升至 4.65， $|t|>2$ 占比从 69.44%和 69.60%提升至 71.27%。但总体 IC 均值绝对值求平均并未有明显提升，首次挖掘 IC 均值绝对值求平均为 0.016，RankIC 均值绝对值求平均为 0.025，第一次优化后略有降低，分别变为 0.013 和 0.023，第二次优化后变为 0.016 和 0.025，EvalGPT 对于表现良好的因子优化效果更为明显。

图表20：第一次优化后因子累计周度 IC



资料来源：Wind, OpenAI, 华泰研究

图表21：第一次优化后因子累计周度 RankIC



资料来源：Wind, OpenAI, 华泰研究



资料来源：Wind，OpenAI，华泰研究

总结

本文是大语言模型与多智能体用于量价因子挖掘的深入实践。在本研究构建的“GPT 因子工厂”中，三类角色各司其职：（1）FactorGPT，负责构建因子表达式；（2）CodeGPT，负责将因子表达式转换为可自动执行的程序，并计算结果；（3）EvalGPT，负责对因子计算结果进行回测检验，同时生成优化建议，并可将建议返回给 FactorGPT，进而继续优化该因子。GPT 因子工厂二次优化后的因子中，IC 表现靠前（前 20%）绝对值均值为 0.034，RankIC 表现靠前（前 20%）绝对值均值为 0.054，因子表现整体较好。

大语言模型的发展如火如荼，大模型应用层出不穷。客观而言，目前大语言模型本身的能力依然尚显单薄，为了进一步发挥大语言模型的能力，使其能自动化地完成高复杂度任务，构建更为复杂且具备影响力的应用，学界持续探索基于大模型的多智能体（Multi-Agent）框架，智能体本身的主动性、反应性等能力大幅扩展大模型的能力边界，而多智能体间的分工合作进一步显示出“群体的智慧”。本文利用大语言模型和多智能体实现量价因子挖掘框架，是大语言模型应用于量化研究的重要探索。

本文的主要结果及结论如下：

1. 基于 GPT 和多智能体实现的因子挖掘框架，可持续挖掘复杂有效的量价因子，并且由于大语言模型输出的随机性，挖掘出的因子间相关性也在可接受范围内。
2. 对于因子挖掘效果，首次因子挖掘获得的因子中，分层 1 年化超额收益均值为 11.14%，分层 5 年化超额收益均值为 -1.11%，分层 1-分层 5 年化超额收益均值为 12.25%，因子分层效果总体较好。
3. 对于因子相关性，首次因子挖掘中，因子相关系数绝对值均值为 0.229，优化后变为 0.192 和 0.230，第一次优化后与第二次优化后的因子相关性水平依旧保持，因子工厂产出的因子相关性普遍较低。
4. 对于因子优化效果，第二次优化后，相较首次挖掘和第一次优化，显示出较为明显的提升效果，例如 $|t|$ 均值从 4.57 和 4.53 提升至 4.65， $|t| > 2$ 占比从 69.44% 和 69.60% 提升至 71.27%。
5. 因子释义和优化建议方面，GPT 相比于传统因子挖掘框架（例如遗传规划）显示出较大优势，词频分析显示 FactorGPT 能够从底层表字段出发，较准确显示出因子的量价特征；EvalGPT 则擅长从窗长等角度对因子提出优化建议。

本文仍有多项未尽之处：（1）本研究测试使用的量价底层表字段较少，可尝试增加字段提升因子构建复杂度；（2）本研究仅涉及量价特征，基本面、一致预期以及宏观中观指标的加入能否提升因子挖掘效果，值得深入探索；（3）本文测试中仅进行两次额外优化，更多次优化后能否持续提升因子效果，也可深入评估。

参考文献

- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... & Wen, J. R. (2023). A survey on large language model based autonomous agents. arXiv preprint arXiv:2308.11432.
- Wang, S., Yuan, H., Zhou, L., Ni, L. M., Shum, H. Y., & Guo, J. (2023). Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment. arXiv preprint arXiv:2308.00016.
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., ... & Hu, X. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond. arXiv preprint arXiv:2304.13712.

风险提示

GPT 挖掘因子是对历史的总结，具有失效风险。GPT 挖掘量价因子可解释性受限，使用需谨慎。大模型训练集广泛，可能存在过拟合风险。

免责声明

分析师声明

本人，林晓明、何康，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师林晓明、何康本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15% 以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15% 以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国：华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J

香港：华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：AOK809

美国：华泰证券（美国）有限公司为美国金融业监管局（FINRA）成员，具有在美国开展经纪交易商业业务的资格，经营业务许可编号为：CRD#:298809/SEC#:8-70231

华泰证券股份有限公司

南京

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码：518017

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com

华泰金融控股（香港）有限公司

香港中环皇后大道中99号中环中心58楼5808-12室

电话：+852-3658-6000/传真：+852-2169-0770

电子邮件：research@htsc.com

http://www.htsc.com.hk

华泰证券（美国）有限公司

美国纽约公园大道280号21楼东（纽约10017）

电话：+212-763-8160/传真：+917-725-9702

电子邮件：Huatai@htsc-us.com

http://www.htsc-us.com

©版权所有2024年华泰证券股份有限公司