

人工智能平台在量化投资中的应用研究

Shuo Wen

三亚大学盛宝金融科技商学院 海南三亚

【摘要】本文旨在对人工智能平台在量化投资领域的应用进行深入的探讨与分析。首先，我们介绍了人工智能平台在量化投资领域的背景以及早期、近期和当前的发展阶段。然后在方法论上，我们以“Qlib”作为典型的人工智能平台案例，说明人工智能平台在量化投资策略开发中的应用影响。为此，我们首先详细梳理了该平台的框架，并对其模块化和流程管理的特点进行了鲜明的阐述，以便更好地理解人工智能平台在量化投资策略设计中的运作机制。然后，我们引入了一种性能评估方法，将 Qlib 与其他传统解决方案进行了比较。结果表明，人工智能平台的应用将有效缩短加载时间，通过利用多核 CPU 的效用使设计过程更加高效，同时由于模块化，开发过程更加灵活。最后，我们讨论了人工智能平台应用的局限性，并展望了未来的发展趋势。

【关键词】人工智能平台应用；Qlib；平台框架；性能评估

【基金项目】Research on the Curriculum Reform of "New Engineering, New Arts, New Medicine and New Business" in quantitative investment major (SYJGSX202259)

【收稿日期】2024 年 10 月 22 日

【出刊日期】2024 年 11 月 20 日

【DOI】10.12208/j.aiml.20240004

Research on the Application of Artificial Intelligence Platform in Quantitative Investment

Shuo Wen

Saxo Fintech Business School, University of Sanya, Sanya, Hainan

【Abstract】The purpose of this article is to provide for an in-depth discussion and examination of the application of AI platform in quantitative investment field. First, we introduce the background of the AI platform in quantitative investment as well as its development stage in early, recent and current time. Then in terms of methodology we take “Qlib” as a typical AI platform case to illustrate the impact of the application of AI platform in the development of quantitative investment strategy. To do it first the framework of the platform was sort out in detail with distinctive characteristics of modularization and process management in order to better understand the operation mechanism of the AI platform in the design of quantitative investment strategy. Then a performance evaluation method is introduced to make the comparison between the Qlib with other traditional solutions. The outcome shows that the application of AI platform will effectively shorten the loading time and make the design process more efficient by leveraging the utility of multi-core CPUs as well as making the development process more flexible due to modularization. Finally, we discuss the limitation of the application of AI platform and look forward to the development trend in the future.

【Keywords】Application of AI platform; Qlib; Platform framework; Performance evaluation

1 简介

1.1 宏观背景

量化投资是一个成熟的领域，它使用数学模型、算法和数据分析来做出投资决策。近年来，人工智

能（AI）的应用作为提高投资决策效率和准确性的手段引起了广泛关注。近年来，人工智能在量化投资中的应用已成为一个越来越受关注的话题^[1]机器学习、深度学习等人工智能技术的快速发展，催生

注：本文于 2023 年发表在 Advances in Computer and Communications 期刊 4 卷 3 期，为其授权翻译版本。

出可用于指导投资决策的复杂算法，有望大幅提高投资决策的效率和准确性，成为量化投资领域从业者和研究人员关注的重要课题。人工智能算法可以实时分析大量数据，识别出可能被人类投资者忽视的模式和关系，还可用于开发预测模型，使投资决策基于前瞻性预测而非过往表现。

虽然关于人工智能在量化投资中的应用的研究越来越多，但这一领域仍需要进一步研究。一个特别令人感兴趣的领域是检查 Qlib 等人工智能平台的性能，因为迄今为止，这在文献中受到的关注相对有限。此外，还需要进一步研究人工智能在投资中的潜在缺点，例如它对数据质量的依赖以及所用算法的可解释性。

1.2 人工智能平台在量化投资领域的应用发展

(1) 人工智能在量化投资中的早期应用：

在量化投资领域人工智能发展的早期阶段，主要关注的是使用人工智能算法来自动化识别潜在投资机会的过程。这是通过使用人工智能算法来分析大量数据（例如金融和经济数据）来识别模式并做出预测来实现的。

量化投资领域使用的早期人工智能算法包括决策树、人工神经网络（ANN）和遗传算法^[2]。这些算法能够处理大量数据，识别模式并做出预测。人工智能算法在量化投资中的应用使投资专业人士能够快速分析市场趋势，识别被低估的资产并做出明智的投资决策。

人工智能在量化投资中早期应用的具体例子包括算法交易、投资组合优化和风险管理。算法交易涉及使用计算机程序根据数学模型和算法执行交易^[3]。投资组合优化是选择最佳资产组合以实现期望的回报水平，同时最小化风险的过程。人工智能算法用于分析财务数据并确定给定投资组合的最佳资产组合。在风险管理中，人工智能算法用于分析市场趋势，识别潜在风险并制定降低这些风险的策略。

(2) 人工智能算法和工具的进步：

近年来，随着人工智能算法和工具的进步，人工智能在量化投资领域的发展也日趋成熟。深度学习算法（如卷积神经网络（CNN）和循环神经网络（RNN））在该领域的应用越来越普遍。这些算法可以分析大量数据，使投资专业人士能够识别出原本会被忽视的模式和关系。这些信息随后可用于指导投资决策，从而做出更明智、更准确的预测。

推动人工智能在量化投资领域增长的另一个因素是云计算平台的发展。这些平台为投资专业人士提供了强大的计算资源，使他们能够更快、更高效地运行复杂的模拟和分析。这促进了人工智能在投资领域的发展，因为投资专业人士现在可以在云端访问实施复杂投资策略所需的计算资源，这更加灵活和方便

(3) 量化投资人工智能平台的出现：

随着人工智能在量化投资领域的发展，许多专门为投资专业人士服务的人工智能平台应运而生。这些平台提供了一系列工具和算法，旨在帮助投资专业人士做出更明智的投资决策。

例如，Qlib 是一个流行的量化投资人工智能平台，它提供了一系列工具和算法，帮助投资专业人士分析数据并做出投资决策^[3]。它具有用户友好的界面，旨在让各个经验水平的投资专业人士都能轻松使用。例如，Qlib 的预测模型可以分析大量数据，帮助投资专业人士预测市场趋势并根据这些趋势做出投资决策。

综上所述，近年来，人工智能平台在量化投资中的应用发展迅速，这得益于计算能力的提高、大数据的增长以及对更复杂投资策略的需求。随着 Qlib 等人工智能平台的出现，投资专业人士现在可以使用强大的工具和算法来帮助他们做出更明智的投资决策。

2 方法论

本文以 Qlib 为典型案例，阐述人工智能平台在量化投资中的应用方法论。Qlib 是量化投资领域的领先人工智能平台，提供一系列用于投资组合优化、风险管理和预测分析的工具和算法。Qlib 使用机器学习算法分析大量财务数据，从而识别出可能被人类投资者忽略的模式和关系，并根据这些信息为投资决策提供参考，优化投资组合绩效。

2.1 Qlib 人工智能平台框架

说明 AI 平台的框架将更容易理解目标与其他平台相比的表现。一般来说，开发量化投资策略的整个过程包括三个步骤：数据预处理、模型开发和绩效分析。每个步骤的开发过程分为几个模块，使研究人员能够更灵活地开发具体项目。建模模块被设计为高度用户可定制的，这样建模设计和量化投资策略的开发将更加动态。下图显示了 Qlib 的整体框架：

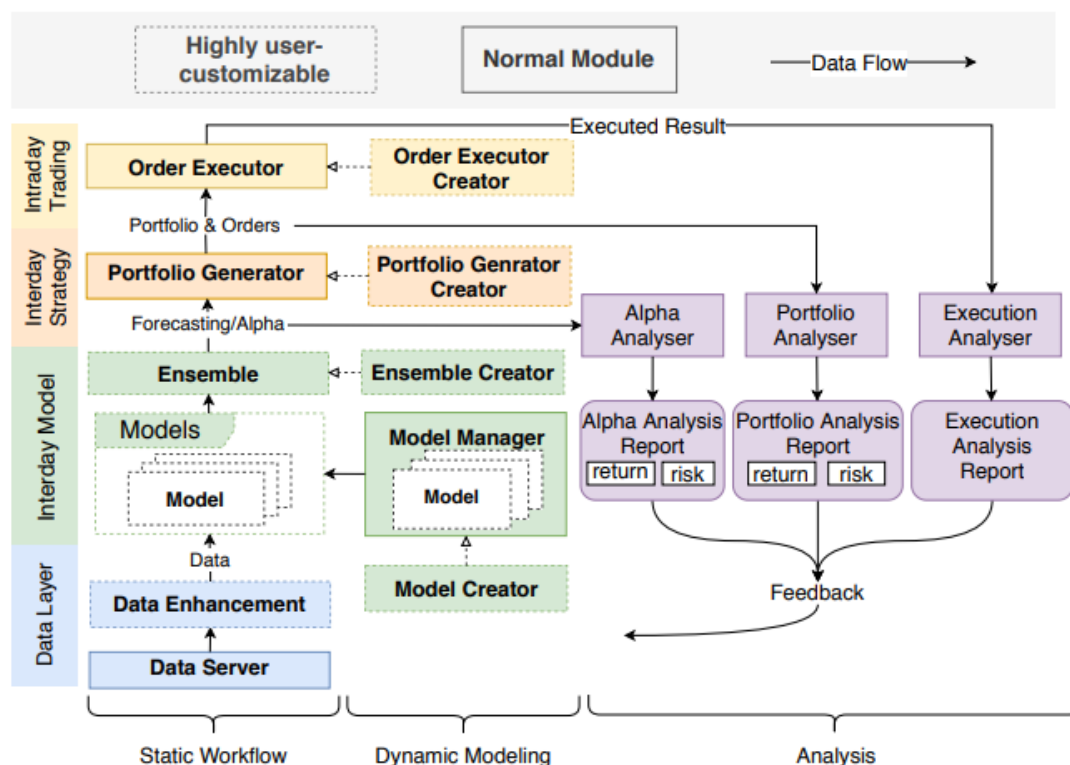


图 1 Qlib 和模块的框架

从上图 1 可以看出，整个系统按照目前现实的研究流程被划分为各个独立的模块。无论是传统的还是基于 AI 的量化投资研究方向，大部分都可以看作是一个或多个模块接口的实现。在每个模块中，Qlib 都提供了一些对用户有用的典型实现。此外，模块还让研究人员可以自由地放弃现有的方法，转而研究新概念。在这样的框架下，研究人员可以用较少的成本通过额外的模块来测试新概念并提高整体性能。

这样的目标导致系统设计上采用模块化设计，系统按照目前现实的研究流程被划分为各个独立的模块。无论是传统的还是基于人工智能的量化投资研究方向，大部分都可以看作是一个或多个模块接口的实现。在每个模块中，Qlib 都提供了一些对用户有用的典型实现。此外，模块还让研究人员可以自由地放弃现有的方法，转而研究新概念。在这样的框架下，研究人员只需在其他模块上投入少量资金，就可以评估新概念和整体表现。

要了解该项目在平台框架中是如何开发的，让我们从左下角的“数据服务器模块”开始，它提供了一个用于请求和处理未处理数据的数据引擎。研究人员可以使用检索到的数据在“数据增强模块”中创建自己的数据集。提供数据集的另一种研究方法是创建用于训练的数据集。模型由“模型创建器模块”使用数据集进行学习。近年来，许多研究人员研究了各种模型，试图从金融数据集中提取交易信号。此外，“模型创建器模块”受益于元学习提供的新颖的学习范式。

模型管理系统已发展成为工作流程的关键组成部分，因为在现代研究工作流程中，有多种方法可以对财务数据进行建模。当代量化研究人员的“模型管理器模块”解决了这些问题。由“模型集成模块”支持的集成学习是一种使用各种模型来提高机器学习模型性能和耐用性的流行方法。“投资组合生成器模块”旨在通过从模型产生的交易信号中创建投资组合来进行投资组合管理。该平台提供高保真交易模

拟器,包括“订单执行器模块”和“分析器模块”,可自动分析交易信号、投资组合表现和执行结果。作为响应式模拟器而非回溯测试功能,“订单执行器模块”可以作为学习范式的基础,该范式要求对“分析器模块”创建的环境进行反馈。

量化投资使用随时间更新的时间序列数据。样本数据集会随着时间的推移而变大。我们经常频繁更新模型,以便利用新数据。由于股票市场的动态性质,动态更新的模型和交易技术除了更好地利用不断增长的样本数据外,还将提高性能。因此,在“静态工作流”中使用一组静态模型和交易技术肯定不是最佳选择。在量化投资中,模型和策略的动态更新是一个关键的研究领域。“动态建模”模块提供支持此类系统所需的接口和架构。

2.2 性能评估

	HDF5	MySQL	MongoDB	InfluxDB	Qlib -E -D	Qlib +E -D	Qlib +E +D
Storage(MB)	287	1,332	911	394	303	802	1,000
Load Data(s)	0.80±0.22	182.5±4.2	70.3±4.9	186.5±1.5	0.95±0.05	4.9±0.07	7.4±0.3
Compute Expr.(s)	179.8±4.4				137.7±7.6	35.3±2.3	-
Convert Index(s)	-				3.6±0.1		-
Filter by Pool(s)	3.39 ±0.24						-
Combine data(s)	1.19±0.30						-
Total (1CPU) (s)	184.4±3.7	365.3±7.5	253.6±6.7	368.2±3.6	147.0±8.8	47.6±1.0	7.4±0.3
Total(64CPUs) (s)	-				8.8±0.6	4.2±0.2	-

图2 不同存储解决方案的性能比较

注意,每个估值指标的含义如下:

①加载数据:由于科学计算采用基于数组的格式,因此将 OHCLV 数据或缓存加载到 RAM 中。

②计算表达式:计算派生因子。

③转换索引:仅适用于 Qlib。由于 Qlib 在原始数据中不存储索引,因此需要设置数据索引。

④过滤数据:按特定池子过滤股票数据。例如,SP500 总共包含 1000 多只股票,但每天只包含 500 只股票。特定日期未包含在 SP500 中的数据应被过滤掉,尽管它曾经出现在 SP500 中。加载数据时无法过滤掉数据,因为一些派生特征依赖于历史 OHLCV 数据。

⑤合并数据:将不同股票的所有数据连接成一个基于数组的数据

从图二可以看出, Qlib 的紧凑存储实现了与专

对于像 AI 技术这样的数据驱动方法论,数据处理的效率至关重要。Qlib 作为以 AI 为中心的平台,提供了数据处理和存储的解决方案。为了展示 Qlib 的性能,我们将其作为 AI 平台的典型示例与其他几种解决方案进行了比较,其中包括 HDF5、MySQL、MongoDB 和 InfluxDb。下文中显示的“Qlib+E-D”指示符表示启用了表达式缓存并禁用了数据集缓存的 Qlib,依此类推。

解决方案的测试方法是从股票市场的基本 OHLCV7 日数据创建数据集,其中包括数据查询和处理。最终数据集来自 OHLCV 数据的 14 个因子组成。数据的时间范围从 1/1/2007 到 1/1/2020。股票池每天由 800 只股票组成,每天都会变化。

3 分析与结果

各平台间性能评估结果如下:

用的科学 HDF5 数据文件相似的大小和加载速度。数据库在数据加载上花费了太多时间。研究底层实现后,我们可以发现,无论是时间序列数据库还是通用数据库解决方案,数据都经过了太多层接口和不必要的格式转换,这大大减慢了数据加载过程。由于 Qlib 的内存缓存,“Qlib-E-D”与 Compute Expr 的指标相比节省了大约 24%的时间。此外, Qlib 提供了表达式缓存和数据集缓存机制。在“Qlib+E-D”中启用表达式缓存后, Compute Expr 的时间节省了 80.4%。如果没有错过表达式缓存,则可以节省。除了计算成本之外,最耗时的步骤是数据组合。将特征组合成每只股票的一份基于数组的数据占了“Qlib+E-D”的大部分时间,而这包含在“计算表达式”步骤中。为了减少这些成本,创建了数据集缓存。时间成本进一步降低,如“Qlib+E-D”列所示。此外,

Qlib 可以通过使用多个 CPU 核心来加快计算速度。图 2 的最后一行显示, 对于具有多个 CPU 的 Qlib, 时间成本显著降低。“Qlib+E+D”无法进一步加速, 因为它只是读取现有缓存, 几乎不计算任何内容。

4 结论

人工智能平台在量化投资中的应用, 有效帮助量化研究人员充分掌握人工智能技术在量化投资模型和策略开发中的巨大潜力, 同时有效缩短开发周期, 减少加载时间, 从而更高效地提高策略执行效率。以 Qlib 为量化投资人工智能平台的典型代表, 相比其他存储解决方案, 使用 Qlib 加载数据耗时更少, 利用多核 CPU 的优势, 数据处理效率更高。最后, 由于独立的模块化设计和流式架构, 量化研究人员可以从繁琐的任务中解放出来, 主要专注于模型设计和策略开发, 与没有人工智能平台支持的传统量化投资项目开发相比, 具有高度的用户可定制性和灵活性。

尽管应用 Qlib 等人工智能平台有可能提高量化投资决策的效率和准确性, 但在使用这些平台时必须考虑一些限制和假设。其中一个限制是平台使用的数据质量。如果数据不准确或不完整, 平台的性能可能会受到影响。

另一个限制是平台使用的算法的复杂性。虽然使用机器学习算法可以改善投资决策, 但这些算法也可能很复杂且难以解释[6]因此, 在使用人工智能平台进行投资决策之前, 必须对算法及其使用的数据有深入的了解。

最后, 还必须考虑平台做出的假设[6]。例如, Qlib 解决的优化问题假设投资组合中资产的未来收益将呈正态分布, 但在现实投资场景中可能并非总是如此。

尽管存在这些局限性, 人工智能平台在量化投资中的应用仍有可能极大地改善投资决策和投资组合绩效。需要进一步研究以更好地了解这些平台的局限性和假设, 并开发提高其绩效的方法。

参考文献

- [1] Klaus Adam, Albert Marcet, and Juan Pablo Nicolini. Stock market volatility and learning, 2016.
- [2] Franklin Allen and Risto Karjalainen. Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2):245-271, 1999.
- [3] John Bollinger. Bollinger on Bollinger bands. McGraw Hill Professional, 2002.
- [4] Kristina Chodorow. MongoDB: the definitive guide: powerful and scalable data storage." O'Reilly Media, Inc.', 2013.
- [5] Yue Deng, FengBao, Youyong Kong, ZhiquanRen, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, 28(3):653-664, 2016.
- [6] FuliFeng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. Enhancing stock movement prediction with adversarial training. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5843-5849. AAAI Press, 2019.
- [7] V Kaul, S Enslin, SA Gross. History of artificial intelligence in medicine. *Gastrointestinal endoscopy*, 2020.
- [8] A Bohr, K Memarzadeh. The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in healthcare*, 2020.