

# 数智赋能，可量得失

——机器学习新技术助推基金行业数字化转型研究

小队名称：顿学累功

队员：

陈翰文	上海交通大学	大一	金融学-计算机科学
徐天奕	上海交通大学	大一	金融学-计算机科学
傅嘉伟	上海交通大学	大一	经济管理试验班
陆冯毅	上海交通大学	大一	经济管理试验班

# 目录

## 一、行业概览

- 1. 基金的发展历程.....1
- 2. 基金的分类.....1
- 3. 基金的运作和监管.....1
- 4. 机器学习技术对基金的影响.....2

## 二、竞争格局

- 1. 基金行业一览.....2
- 2. 数字化浪潮下，基金行业在量化技术上的竞争格局.....4
- 3. 竞争格局的新特点.....6

## 三、行业表现

- 1. 基金行业.....7
- 2. 机器学习技术市场运用.....9
- 3. 机器学习技术的不足.....10

## 四、行业展望

- 1. 机器学习新范式有望被应用，前沿技术的基金领域水平创新.....12
- 2. 双重机器学习新技术有望在基金应用场景下减少模型误差.....12
- 3. 将大数据纳入人类语境，语言模型将窥测超越数字的市场走向.....13
- 4. 金融改革深化增加量化产品配置，机器学习有充分应用需求.....14
- 5. 政策导向推动技术发展赋能金融场景.....14

## 五、风险管理

- 1. 机器学习在变量间因果识别方面面对复杂技术挑战.....14
- 2. 机器学习的新方法受到数据来源的制约.....15
- 3. 基金经理和投资方思维与行为习惯加大机器学习使用推广难度.....15
- 4. 机器学习技术发展初期对人才提出高要求.....16

## 附录

- 参考文献.....17

## 一、行业概览

基金是一种将多个投资者的资金汇集起来，由专业的基金管理人进行投资和管理的一种集合投资方式。基金可以分为公募基金和私募基金两大类。公募基金是向不特定投资者公开发行的证券投资基金，其发行、管理、托管、销售等环节都受到严格的法律法规和监管部门的规范。私募基金是向特定投资者非公开发行的证券投资基金，其发行、管理、托管、销售等环节相对灵活，但也要遵守相关的法律法规和行业自律规范。基金以其专业化管理、风险分散化、低门槛等优势，受到了广大投资者的青睐。随着机器学习技术的发展，基金也在探索利用人工智能提升投资能力和运营效率的可能性。下面简要介绍基金的发展历程，基金的分类，基金是的运作和监管，机器学习技术对基金的影响几个方面。

### （一）基金的发展历程

基金作为一种集合投资制度，起源于欧洲，发展于美国，传播于全球。基金的发展历程可以分为以下几个阶段：

#### 1. 萌芽阶段

基金的雏形可以追溯到 18 世纪中期的荷兰，当时有一些投资者联合起来，共同出资购买政府债券或股票，并按照出资比例分配收益和风险。这种形式被称为“投资信托”（Investment Trust），是最早的集合投资方式。

#### 2. 成长阶段

基金的成长阶段主要发生在 19 世纪末到 20 世纪中期的美国，这一时期出现了多种类型的基金，如封闭式基金、开放式基金、指数基金、货币市场基金等，并逐渐形成了完善的法律制度和监管机制。美国也成为了全球最大的基金市场。

#### 3. 发展阶段

基金的发展阶段主要发生在 20 世纪下半叶到 21 世纪初的全球范围内，这一时期基金行业呈现出多元化、国际化、创新化的特征，出现了许多新兴的基金品种和市场，如对冲基金、私募股权基金、养老基金、ETF、FOF 等，并涉及到多种资产类别和投资策略。

#### 4. 转型阶段

基金的转型阶段主要发生在 21 世纪以来，这一时期受到全球金融危机、科技革命、社会责任等因素的影响，基金行业面临着新的机遇和挑战，需要进行数字化转型、智能化升级、绿色化改革等方面的调整和创新，以适应变化的市场环境和客户需求。

### （二）基金的分类

按照投资标的的不同，基金可以分为股票型基金、债券型基金、货币市场型基金、混合型基金、指数型基金、FOF（Fund of Funds）型基金等多种类型。其中，股票型基金主要投资于股票市场，追求长期增值；债券型基金主要投资于债券市场，追求稳定收益；货币市场型基金主要投资于货币市场，追求流动性和安全性；混合型基金可以投资于多种市场，平衡收益和风险；指数型基金主要跟踪某一指数的表现，追求低成本和高效率；FOF 型基金主要投资于其他基金，实现多层次的分散化。

按照对机器的使用程度的不同，基金可以分为传统型基金、智能型基金和全自动型基金三种类型。其中，传统型基金主要依靠人工进行投资和管理，机器只起到辅助作用；智能型基金主要利用机器学习技术进行数据分析和策略生成，人工只起到监督和调整作用；全自动型基金主要由机器完成投资和管理的全过程，人工只起到监管和评估作用。

### （三）基金的运作和监管

基金的运作和监管涉及多个机构和部门。其中，最重要的是各国的证券监管机构，如美国的证券交易委员会（SEC）、欧盟的欧洲证券与市场管理局（ESMA）、中国的中国证券监督管理委员会（CSRC）等。这些机构负责制定和执行有关基金市场的法律法规和政策，对基金的发行、管理、托管、销售等环节进行审批、登记、备案、检查、处罚等监管措施。

除了证券监管机构外，还有其他机构参与到基金运作和监管中来。例如，中央银行负责监管银行和信托等机构的运作；证券交易所负责对基金的交易行为进行监控和规范；行业协会负责行业的自律管理职能，如组织从业人员资格考试、发布行业报告等；基金评价机构负责为基金提供评级和评价服务，如晨星、标普等，它们通过对基金的业绩、风险、风格等方面进行分析和比较，为投资者提供参考信息。

### （四）机器学习技术对基金的影响

机器学习技术是一类应用大量统计学理论，从数据中自动分析获得规律，再利用规律对未知数据进行预测的算法。机器学习技术已广泛应用于数据处理和数据挖掘、自然语言处理、证券市场分析、检测交易欺诈等领域。基金也在探索利用机器学习技术提升投资能力和运营效率的可能性。

在基金投研领域，机器学习技术可以帮助研究人员处理和分析海量的金融数据，从中挖掘出有价值的信息和模式，为投资决策提供支持。例如，机器学习技术可以通过对历史数据的回测，评估不同的投资策略的有效性和稳健性；可以通过对新闻、社交媒体、企业报告等文本数据的自然语言处理，提取出对股票价格有影响的事件和情绪；可以通过对市场行情和交易数据的实时监测，发现异常波动和潜在的套利机会等。

在基金运营领域，机器学习技术可以帮助运营人员提高工作效率和降低风险。例如，机器学习技术可以通过对基金持仓和交易数据的分析，识别出可能存在的错误或违规操作，并及时发出预警或纠正；可以通过对基金净值和收益率等指标的预测，优化基金估值和定价过程；可以通过对基金风险指标和风险因子的计算，监控基金的风险状况，并提供风险管理建议等。

在基金营销与客户服务领域，机器学习技术可以帮助营销人员和客服人员提高服务质量和客户满意度。例如，机器学习技术可以通过对客户行为数据的分析，构建客户画像，并根据客户特征和需求推荐合适的基金产品；可以通过对客户反馈数据的分析，评估客户满意度，并根据客户意见改进服务流程；可以通过对常见问题和答案的自然语言处理，构建智能问答系统，并实现自动化或半自动化的客服服务等。

## 二、竞争格局

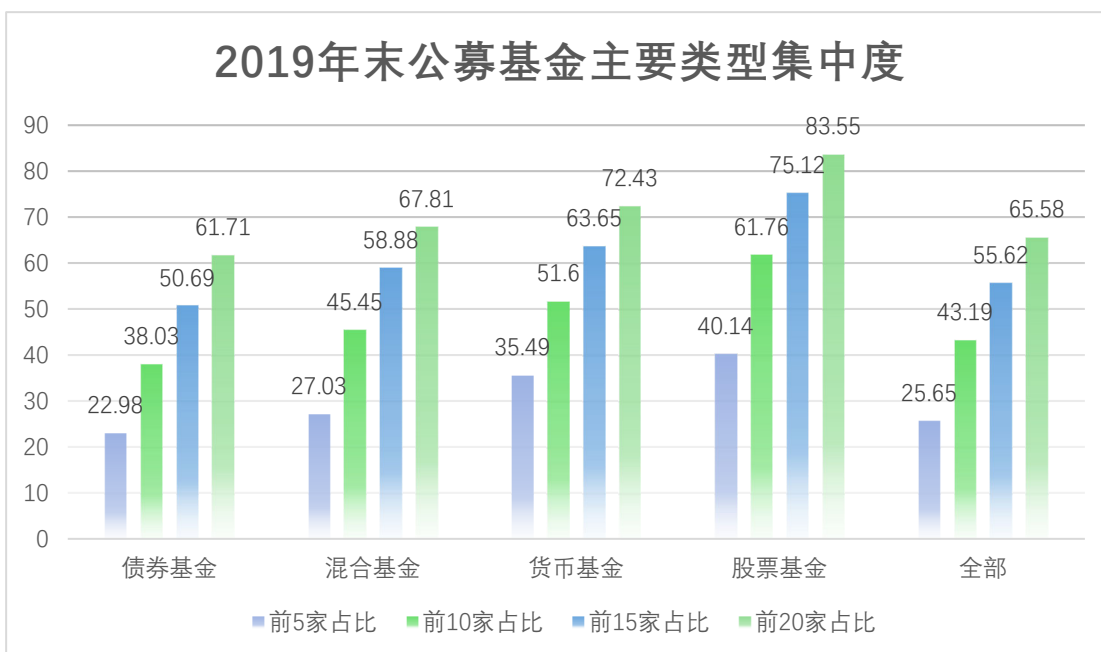
### （一）基金行业一览

#### 1. 市场集中度：

由下图可以分析得出：2019年公募基金行业市场集中度整体处于中等水平，头部企业市场规模较大，但远不足以形成支配优势；行业的进入壁垒不大，但随着集中度

提高，需要的市场份额迅速增大，新进入者短期难以形成有效威胁；前期进入市场的老牌基金公司已经开始呈现出一定的规模优势，集中度有进一步提高的空间和可能；但是客户和消费者的议价能力依旧较高，选择较为丰富，个别企业不能完全支配市场。

其中，债券基金和混合基金相对而言市场集中度较小，货币基金和股票基金市场集中度较大；其中股票基金的 CR5 和 CR20 反映出大型股票基金的价格支配能力极强，投资者的议价能力较弱。

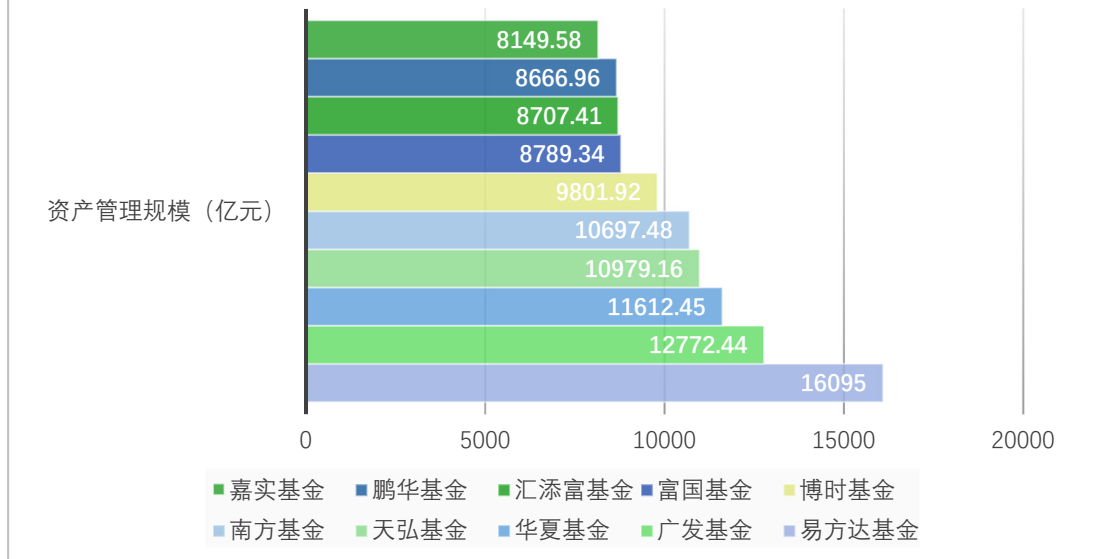


（资料来源：中国证监会、中国证券投资基金业协会）

## 2. 头部基金一览：

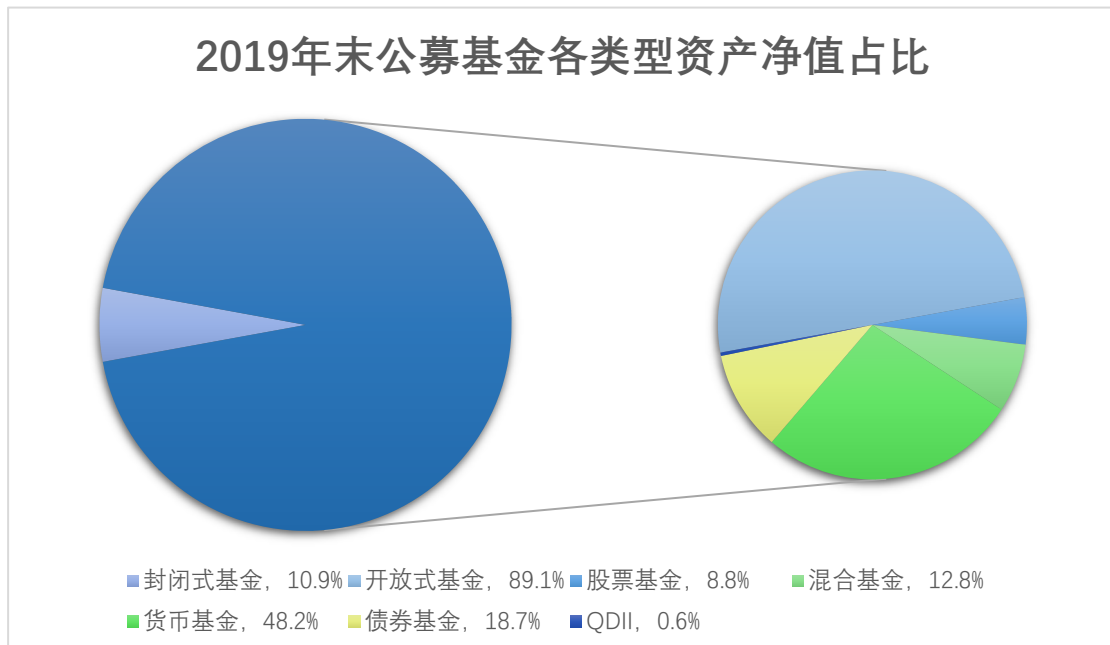
头部基金公司中，易方达基金管理有限公司的资金规模有明显的领先，剩余 9 家基金公司资金规模差异较小，基本维持在 10000 亿元左右。

## 2023年上半年度资金管理规模排名前10的基金公司



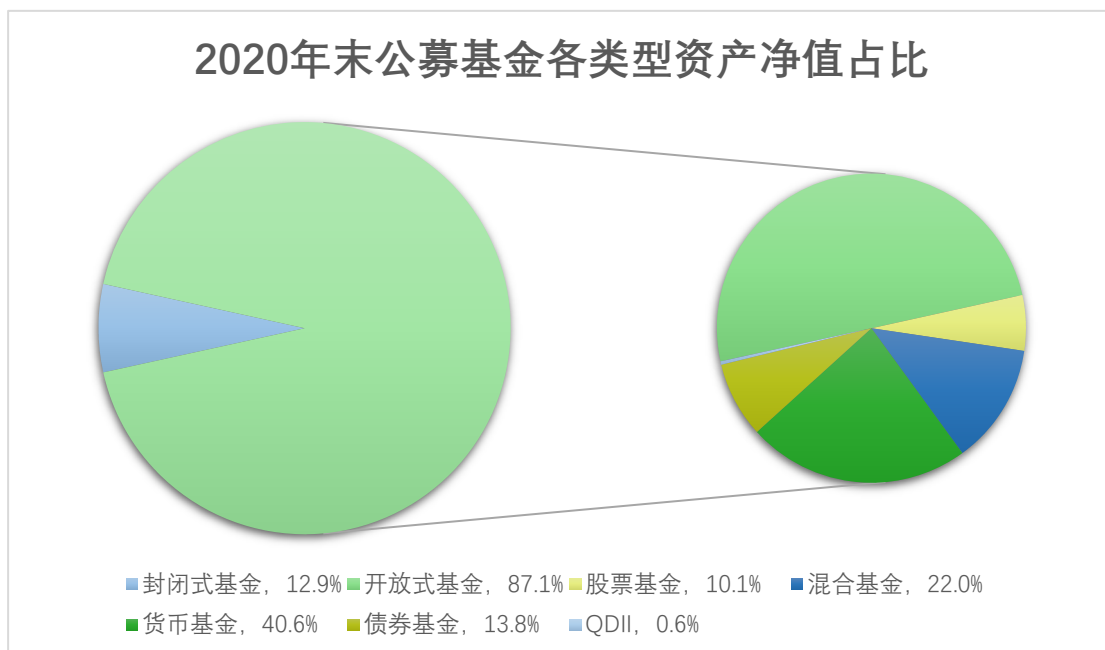
### 3. 不同类型基金产品的市占率：

2019 年末的数据表明：公募基金内的开放式基金中，货币基金占比达到将近五成；而 2020 年末，货币基金和债券基金资产净值占比都小幅下降，混合基金占比大幅上升。





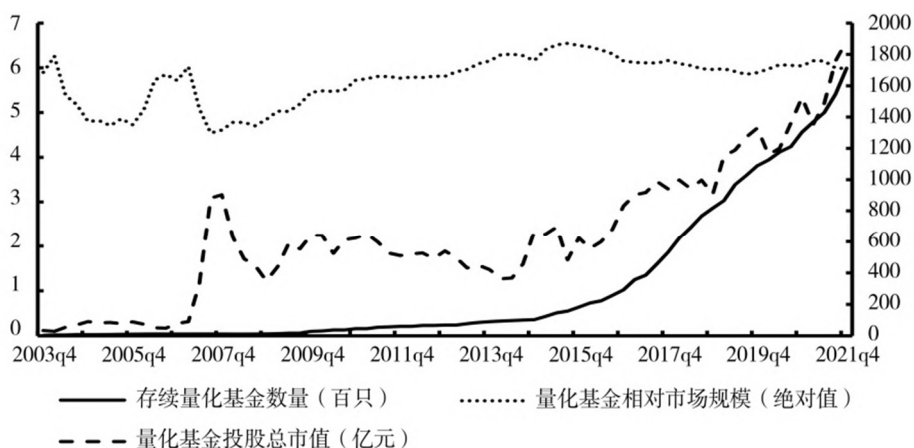
### 2020年末公募基金各类型资产净值占比



## (二) 数字化浪潮下，基金行业在量化技术上的竞争格局

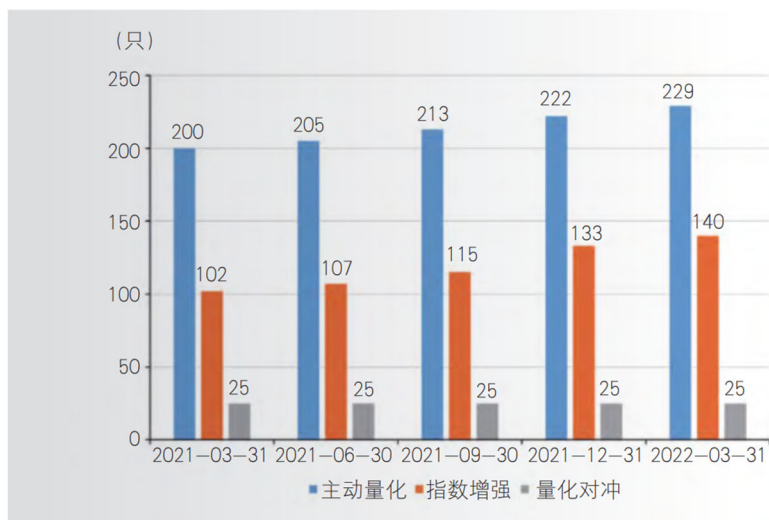
### 1. 随着大数据、人工智能技术发展，近年来竞争愈发激烈

纵向来看，2015年之前，量化基金总数较少，投股总市值基本维持不变。2015年至2021年间，量化基金数量和投股总市值都大幅增加，但相对市场规模绝对值并未有明显变化，侧面反映出量化基金内部竞争之激烈。



(2003年第四季度至2021年第四季度国内量化基金发展示意图)

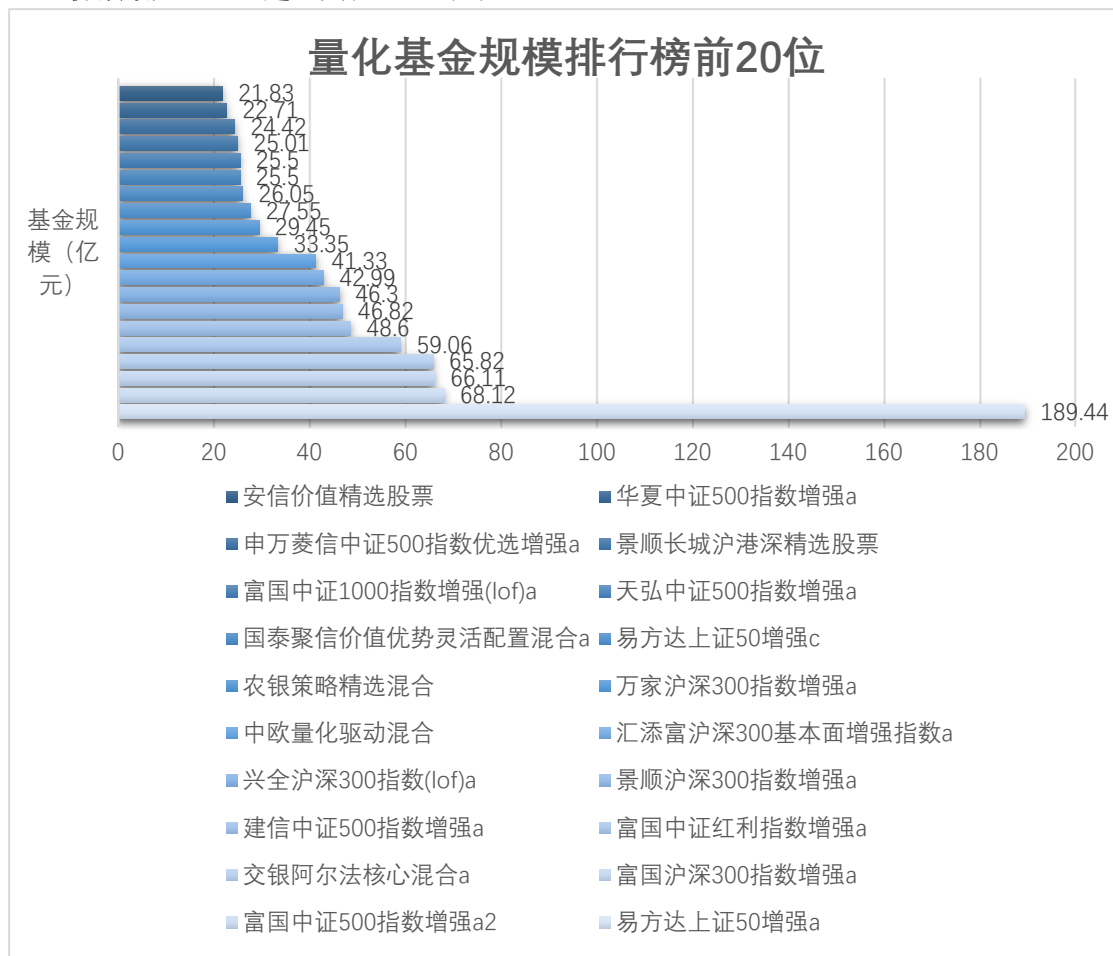
仅2021年一年，主动量化基金增加了29只，指数增强基金增加了38只，增幅分别为14.5%和37.3%。相较于往年数据来看，不难发现，近几年是量化型基金数量的攀升期，其中指数增强型基金增幅最快，竞争最为激烈。



(2021 年第一季度末至 2022 年第一季度末量化型基金数量变化, 数据来源 wind)

## 2. 竞争主要集中在大型基金公司中:

横向来看, 以 2023 年 4 月数据, 从基金规模来看, 量化基金的竞争仍旧是集中在资金规模排名前 10 的基金公司当中。其中, 易方达基金和富国基金所占到的市场份额最大, 远远超过其他基金公司。



(2023 年 4 月 21 日 问财网数据)



观察年收益率排行靠前的量化基金，绝大多数同样都来自于富国、人保、易方达、银华、华宝、长信、博时、华夏、中银、安信等大型基金公司。

### （三）竞争格局的新特点

#### 1. 技术驱动：

人工智能、大数据、云计算等新技术广泛运用，改变着基金销售、投资管理和交易方式。量化投资手法和 ETF 发展迅速，机构投资者利用技术应用量化模型进行高频交易。这使得利用好科技成为基金公司的重要竞争力。

#### 2. 产品创新：

各公司通过发行新产品来满足不同投资者的需求，如 ETF、量化对冲、指数增强等。产品策略的新颖性和特色会影响公司的竞争优势。

#### 3. 个性化服务：

各基金公司提高客户服务水平，利用大数据提供个性化的投资解决方案和资讯服务。这也成为判断公司竞争力的重要维度。

#### 4. 社会责任：

ESG 投资理念兴起，投资者越来越重视基金公司在环境、社会与公司治理方面的责任表现。这使基金公司需要在盈利之外，兼顾社会责任来满足投资者需求。

#### 5. 全球化发展：

海外合作发展加剧，跨境基金销售环境优化。部分公司开始布局海外子公司和并购，在全球范围内共享技术、投研和产品。基金行业的竞争正朝全球化方向演进，不同公司的全球化战略和布局能力影响着其竞争优势。

## 三、行业表现

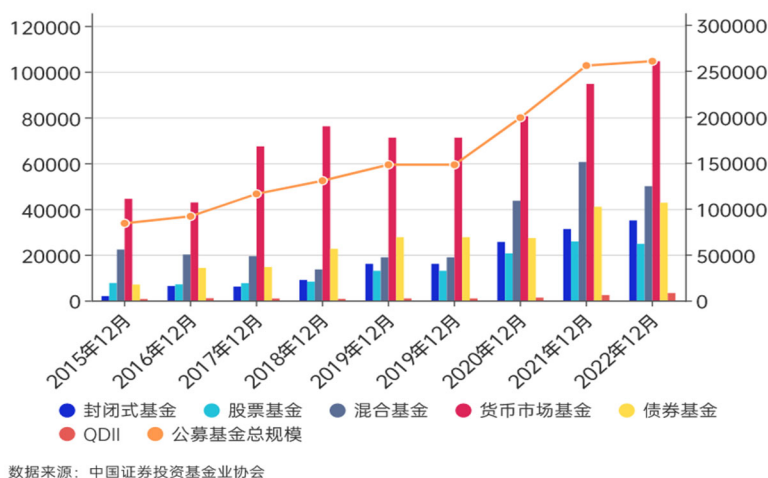
### （一）基金行业

#### 1. 整体市场表现：

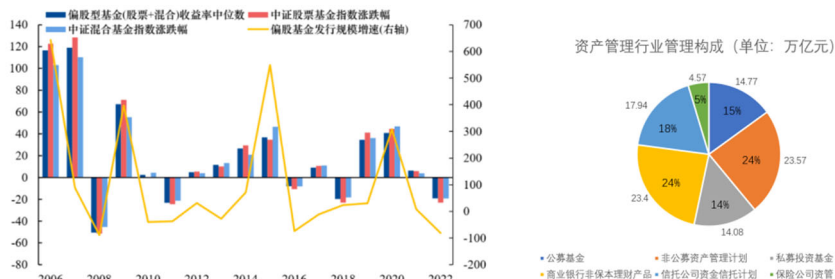
过去几年全球基金规模持续呈现扩大趋势，数据显示 2019 年全球基金规模达到 48.2 万亿美元，较 2014 年数据增长超过 40%，其中美国占据最大规模，占全球比重超 40%；亚洲市场基金规模增长速度超过欧美，特别是中国市场的增速远超世界平均水平。2015 年至 2022 年间中国公募基金总规模 83971.83 亿元增长至 260311.89 亿元，增幅超过 209%。（如图）除了中国以外，其他亚洲国家如日本、韩国等基金规模同样稳步增长。

## 近年国内公募基金规模增长趋势

单位：亿元



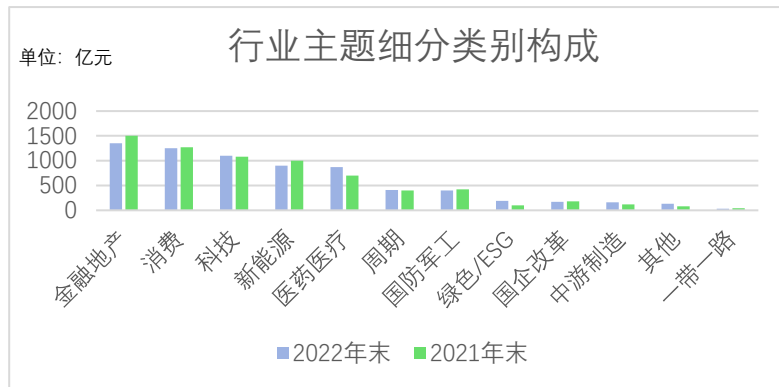
但 2022 年公募基金总体表现不佳，其收益均值为-10.96%，且收益分化大，收益最小值为-70.40%，收益最大值为 68.48%。



## 2. 市场投资方向转变，机器学习足担大任：

### (1) 传统基金类型内部投资领域向科技等新兴产业转移：

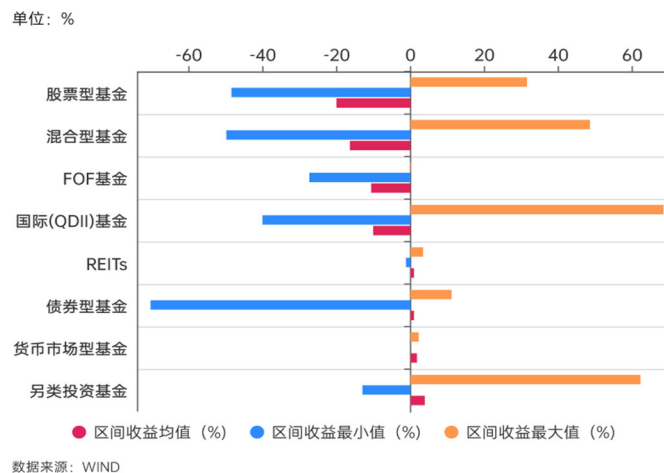
随着科技的迅速发展和新兴产业的不断涌现，基金公司投资方向转向全新的投资领域。例如，科技类基金、新能源基金、医疗健康基金等产品。wind 数据显示在 2016 年至 2019 年的中国基金市场中，科技类基金的规模呈现持续上升趋势，由 2016 年末的约 2.5 万亿元人民币增长至 2019 年末的约 5.5 万亿元人民币，增幅约 120%。其中，在 2018 年第四季度至 2019 年第二季度间，科技类基金规模增长较快，增幅接近 30%，截至 2022 年 3 月底，中国 A 股市场中的科技类基金的规模已经超过了 8000 亿元人民币，与 2015 年相比增长超过 50 倍；新能源基金的规模也呈现持续上升趋势，由 2016 年末的约 3000 亿元人民币增长至 2019 年末的约 1.2 万亿元人民币，增幅约 300%。其中，2017 年和 2018 年是新能源基金规模增长最为迅速的两年，增幅均超过 100%；在 2016 年至 2019 年期间，医疗健康基金的规模也呈现持续上升趋势，由 2016 年末的约 1.6 万亿元人民币增长至 2019 年末的约 3.3 万亿元人民币，增幅约 100%。其中，2018 年是医疗健康基金规模增长最为迅速的一年，增幅接近 60%。



### (2) 另类投资基金回报可观，超过基准预期：

另类投资基金是指公开市场交易的权益资产、固定收益类资产和货币类资产等传统投资类型之外的基金。过去几十年中，传统类型基金是基金市场的主流产品，但近几年对冲基金等另类投资基金崭露头角，占据越来越多的市场占比。在 2022 年公募基金总体表现不尽人意的情况下，另类投资基金收益表现最好，其收益均值为 3.83%，超过了排在第二位货币资产基金 1.72% 的收益均值。

### 2022年公募基金各基金类型业绩表现

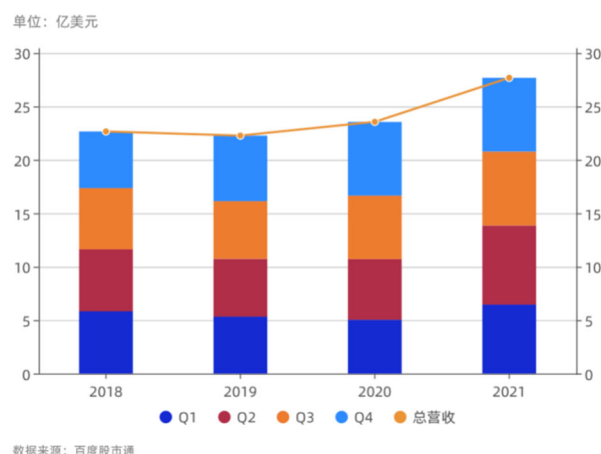


另类投资基金具备更高的投资难度。基金经理需要投资于非传统的资产类别，而这些资产如不动产、商品期货等通常具有高度的不确定性和不稳定性，并且通常没有像传统股票和债券市场那样的流动性和透明度。此外，另类投资基金的准入门槛和评估标准通常也比传统基金更高，投资者需要具备更多的专业知识和经验才可能取得成功。

### (3) 基金市场从被动投资向主动投资转变：

传统上，基金市场以被动投资为主，即跟踪某一指数或行业进行投资，以期望在长期中获得一定的收益。但是，随着市场的变化和投资者的需求，基金公司也开始转向主动投资，即通过研究和分析市场来挖掘投资机会，以获得超过超越市场业绩的收益。资管巨头 Janus Henderson Group 公司在 2017 年宣布计划将其被动投资业务减少一半，并将资源转向主动投资。公司在 2018 年之后的财年中实现了增长，其中主动投资业务的管理费用净收入增长了 18%。Invesco Ltd. 公司在 2018 年花费 44 亿美元收购了 OppenheimerFunds Inc.，旨在增强其在主动投资领域的实力 Inc.。即使是以被动基金闻名的基金公司 Vanguard 也开始扩大其主动基金的规模。该公司旗下的 Vanguard Wellington 基金和 Vanguard Windsor II 基金都是主动管理基金。

Janus Henderson Group PLC年化总营收



综合以上三点可见，基金市场本身作为一个错综复杂的金融市场，正在向更高风险性和复杂性演变。难以预测的投资方向变化以及基金市场向主动投资靠拢的趋势意味着对基金经理在数据挖掘和分析方面更加严苛的要求。基于机器学习的算法分析通过解析大量数据来发掘潜在的投资机会和风险，而这无可以大大提高基金经理的决策效率和投资成功率。

## （二）机器学习技术市场应用

AI 算法正在开启一个全新世代，它帮助基金经理更好地理解市场趋势和股票价格波动，从而进行更加精准的投资决策。目前已经有大量基金公司将机器学习应用于其投资业务。

机器学习技术具备高度适用性。部分基金公司用以分析公司财报、行业报告等信息，从而更好地理解企业的业务情况和前景，以便做出更好的投资决策。它还可以应用于基金组合管理，帮助基金经理优化投资组合，降低投资风险并实现更好的回报。机器学习技术可以为基金经理提供更好的数据分析和决策支持，从而帮助他们取得更好的投资结果。此外，深度学习模型可以对基金经理的投资决策进行风险控制 and 后续监督，确保基金的长期盈利能力和稳定性。总之，无论从因子挖掘、基金组合管理，还是风险优化等方面，机器学习都取得了良好表现。

Bridgewater Associates 公司自主研发 Alpha Engine 系统，该系统使用机器学习和人工智能技术来分析大量的市场数据，帮助公司做出投资决策和管理投资组合。Alpha Engine 系统利用大数据和机器学习技术来对全球市场的复杂关系进行分析，识别出市场趋势和风险，并且自动执行交易策略。

对冲基金公司 Renaissance Technologies 在量化交易领域具有很高的声誉，它利用机器学习和人工智能技术进行投资决策和交易。Renaissance Technologies 开发了一系列的量化交易策略，这些策略利用机器学习和大数据分析技术来识别市场的趋势和规律，从而自动执行交易。该公司的 Medallion 旗舰基金是业界公认的最成功的量化交易基金之一，长期以来都取得了非常好的回报率。另一只对外公布的基金 Renaissance Institutional Equities Fund 在近几年收益相对较差，但曾在很长一段时期内表现出色。

2022年全球十大对冲基金的市场表现

#	公司名字	2022	2021	涨跌幅
1	Bridgewater Associates	\$126,400	\$105,700	19.6%
2	Man Group	\$73,500	\$63,400	15.9%
3	Renaissance Technologies	\$57,000	\$58,000	-1.7%
4	Millennium Management	\$54,968	\$52,314	5.1%
5	Citadel	\$52,970	\$37,650	40.8%
6	D.E. Shaw Group	\$47,861	\$39,738	20.4%
7	Two Sigma Investments/Advisers	\$40,969	\$39,550	3.6%
8	Davidson Kempner Capital Management	\$37,450	\$37,350	0.3%
9	Farallon Capital Management	\$37,400	\$38,100	-1.8%
10	TCI Fund Management	\$36,200	\$40,000	-9.5%

国内采取量化投资策略的大体量基金公司在 2021 年公布数据同样拥有可观收益率  
国内基金年收益率：

数据来源：Choice 终端

公司简称	2021 年收益率 (%)	是否 涉及人工 智能量化	策略
鸣石投资	39.88	是	CTA 策略、指数增强、中性策略
天演资本	37.88	是	——
世纪前沿资产	36.08		对冲指数增强
金戈量锐	33.42		对冲指数增强
佳期投资	31.25	是	人工智能量化
因诺资产	30.53		套利、择时、多因子、CTA
启林投资	24.13		对冲、指数增强
宁波万方量化	20.74	是	——
灵均投资	18.96		中性、多策略、指增
九坤投资	18.22	是	指增、对冲、CTA、多空
幻方量化	18.19	是	——
衍复投资	17.4		中性策略
盛泉恒元	14.82		量化套利、价值投资
诚奇资产	13.48	是	阿尔法
进化论资产	13.06		主观、指增、对冲、多空、FOF
明波投资	12.44		——
金得资产	11.11		对冲
黑翼资产	9.38		股票对冲（指数增强、中性）
千象资产	5.33		CTA、指数增强、复合、量化选股
呈瑞投资	4.85		宏观对冲、CTA、主动量化

机器学习技术在金融量化选股领域的有效性已经在市场表现中逐步得到验证，科技赋能正在助力基金公司追逐超越市场基准的收益。

### （三）机器学习技术的不足

根据已有对量化基金的研究，我们发现量化基金的数量是在近五年呈现一个较为明显的增长的。故我们选取了 20 个 2018 年以前成立的量化基金，考察它们在特定行情下的相关表现。这 20 只基金其中 10 只是非量化混合型基金，10 只为量化基金。这 20 只基金都是从同类型基金中根据资产规模分层抽样选出的。

根据我们预先猜测：量化基金和非量化基金在应对大盘大幅变动的行情时会有更加明显的差异，且量化基金可能有更强的稳定性，随大盘指数变动较小。故我们选择考察了 2018 年以来，上证指数呈现明显的“峰”和“谷”的五段时间（下称转折期），考察指数转折点两侧总时长为 3 个月的 5 个时期。我们在每段时期中通过记录下上证指数和各只基金的周线，分别记录下他们的周涨跌幅。

接着我们分别将每只基金在 5 个转折期内的 11 组涨跌幅值与上证指数的涨跌幅值进行相关性处理，求出他们的相关系数（记为  $\rho_{F_iS}$ ）。接着根据每只基金的资产规模（记为  $FAS$ ）的自然对数进行加权，得出这 10 只量化基金和 10 只非量化基金对于上证指数的 2 个整体相关系数（记为  $\rho_A$ ）。我们近似认为，这 2 个相关系数可以一定程度上近似描述量化基金与非量化基金对于上证指数的相关程度。

计算公式如下：

$$\rho_A = \frac{\sum_{i=1}^{11} \rho_{F_iS} \ln(FAS)}{\sum_{i=1}^{11} \ln(FAS)}$$

根据以上方法，我们计算出的数据如下：

$\rho_A$	拐点 1(谷)	拐点 2(峰)	拐点 3(峰)	拐点 4(谷)	拐点 5(谷)
量化	0.862	0.929	0.816	0.817	0.868
非量化	0.79	0.767	0.764	0.568	0.734

从中数据可以看出，量化基金与非量化基金相比，在行情发生重大转变时其收益率与上证指数的变化的相关系数要比非量化基金平均高 0.134。而其中在拐点 4，即 2022 年 3 月 11 日到 2022 年 5 月 20 日之间的 11 周时间内，量化基金的相关系数要比非量化基金高 0.25 左右。我们可以从中得到两个结论：

#### 1. 机器学习模型在极端行情中的抗风险能力不足

更高的相关系数说明了在选取时间段中量化基金与大盘的关系更大，而我们选取的时间段均为行情剧变，大盘剧烈波动时间，这可以说明量化基金在极端行情中的抗风险能力不如非量化基金。我们分析出可能的原因为：量化基金在面对行情剧变的时期，没有及时的大量调整持仓，而是依据机器学习得出的模型进行较小的持仓调整，这可以认为时机器学习方法在市场预测中的不完全适用，也可以认为目前市场的数据量不足，所训练出的模型不够完善。

#### 2. 机器学习模型对消息面的把控不足

在拐点 4 所在的时间段内，美联储在 2022 年 3 月 16 日宣布加息，将联邦基金利率目标区间上调 25 个基点到 0.75% 至 1.00% 之间，美国货币政策的收紧，导致了全球市场的恐慌性抛售，中国股市也大幅下跌。在此次由消息面主导的下跌中，量化基金的抗风险能力远不如非量化基金。我们分析出的原因可能为：量化基金所用的机器学习模型没有考虑到消息面的重要性，而是一味的依赖市场数据，因此在行情中表现不佳。



## 四、行业展望

### （一）机器学习新范式有望被应用，前沿技术的基金领域水平创新：

目前机器学习的主流范式是孤立学习，既通过给定相关训练集，让算法从中训练得出模型。该学习范式目前广泛应用于各领域，但它的缺点是只能应对相对小范围的有限任务，并且需要人工投喂大量训练数据。事实上，此类范式在面对数据源源不断更新变动的应用场景作用疲软。基金市场随时间演进不断产生新类型、新领域基金，金融大环境也往往随时更迭，新的环境因子也正在不断涌现。通过训练集训练出的学习模型难以自主调整算法模型，进而导致其在投资策略上难以避免的偏差。因此，我们认为在未来另一种学习范式——终身学习（Life-Long Learning）有望应用于基金领域，其前沿技术——类别增量学习算法将在基金组合策略等应用场景拥有良好的发展前景。

终身学习也称为持续性学习，它赋予机器具备面对新环境和新任务自主学习的能力。然而，机器学习模型可能会面临灾难性遗忘问题，即在学习新知识后，可能会忘记之前训练得出的知识结论。灾难性遗忘在传统机器学习范式下难以被避免，而一旦在基金决策上发生将会造成难以挽回的巨额损失。在机器学习前沿领域，类别增量学习算法被提出以解决这个问题，该算法要求模型在完成所有训练过的类别之后进行评估，一个优秀的类别增量模型能够既学习新类知识，又不会遗忘旧类知识。该算法帮助模型模型尽可能适应新类别的特征，而这与瞬息演变的基金市场十分契合。我们相信将类别增量学习算法向基金领域拓展，能够帮助基金经理自主学习新的投资类别和策略，同时保留先前的投资决策和经验，避免机器学习模型灾难性遗忘造成的巨大投资风险。

### （二）双重机器学习新技术有望在基金应用场景下减少模型误差

双重机器学习模型（DML）由 Chernozhukov et al 在 2018 年提出，能对高维的潜在协变量进行控制，并且放松了线性关系的假设，简化了对变量预先协定的过程的同时保证了处理效应技术的准确性。

在金融趋向于国际化的应用场景下，基金公司面临的数据量日益膨胀，这可能使传统方法在处理高位变量时可能面临的准确性问题，例如维度诅咒和双重共线性，或者对于非线性变量间的数据处理难以避免模型设计上的偏误问题等缺陷变得显著。而双重机器学习支持通过诸多机器学习的正则化算法，能自动筛选和预测更高精度、高有效性的控制变量集合，有效缓解了控制变量冗余或控制变量有限导致的估计有偏问题。与此同时，双重机器学习不仅拥有一般机器学习在处理非线性数据上的优势，也能够基于工具变量函数、两阶段预测残差回归以及样本分割拟合的思想，在小样本的情况下保证处置系数估计量的无偏性。因此，双重机器学习的新技术在基金应用领域具有技术上的优越性。随着经济转型增长，该技术的应用范围和需求同时得以扩大，能够满足基金公司在规模扩大时亟须跟进的技术需求。

当前，双重机器学习技术在国内研究尚浅，规模性成果较低，对于该技术的实证应用的研究较多。可见该技术的理论发展仍具有相当潜力，而随着理论发展和突破，其在具体领域的场景应用能力和规模有望得到带动。因此，我们相信双重机器学习技术能提高基金行业决策的准确性，与其在技术上达成互相推进的良性循环。

### （三）将大数据纳入人类语境，语言模型将窥测超越数字的市场走向

过去一年，大型语言模型 ChatGPT 惊艳登场，它所标新立异的一点即是将人类语言纳入大数据范畴，经过大体量文本训练得以获得超越传统机器学习模型的智能。通过转换模型（Transformer），语言模型可以将人类文本转化为向量形式的数值数据，进而实现机器理解人类自然语言的可能。我们认为未来量化基金的机器学习模型若是将语言模型纳入考量范围，有机会解决机器学习模型对市场数据过于依赖的问题，提升行情剧变时的抗风险能力。

目前量化投资所采用的机器学习模型仍停留在基金市场数值分析上，在本文的数据分析环节中同样发现机器学习对大盘指数依赖度更高，而无法敏锐发现其他扩展信息并作出判断。市场波动由多因子共同影响，新闻周刊、股评等自然语言数据同样会影响投资者的投资策略，进而影响整个行业情况。以 2019 年蔓延持续数月之久的澳洲大火为例，当地煤矿、铝土矿等产量受限，遥居北半球的中国能源领域同样受其影响。这类包含突发数据的信息往往第一时间无法以统计数值呈现，而是以自然语言的形式流通。目前基金量化投资模型无法兼顾这一方面，可能会导致决策错误和相应投资风险。我们相信将语言模型应用至该领域可以帮助量化投资以超越数字拘束的视角看待市场走向，获得更好的市场绩效。

#### （四）金融改革深化增加量化产品配置，机器学习有充分应用需求

数据统计显示，截至 2019 年，在大约 465 万亿元的中国居民总资产规模中，金融资产占 147 万亿元，仅包括汽车和住房的实物资产占 318 万亿元。前者人均 10.5 万元，后者人均 22.7 万元。可见当前中国居民部门拥有的总资产规模仍有实物资产比重过高，金融资产比重较低的结构性问题。而随着金融改革的深化、金融部门监管的加强，金融资产和可支配收入比值的上升，资产结构面临优化，使家庭资产结构多元化，增加量化产品的配置。同时，量化产品的扩张也面临着全球化进程的到来。而为了匹配更丰富的资金来源渠道、处理更庞大的数据量，传统的分析手段遭遇挑战，机器学习面临着迫切的应用需求。

#### （五）政策导向推动技术发展赋能金融场景：

《人工智能高水平应用促进经济高质量发展的指导意见》提出了新技术的应用应以场景创新为导向，完成行业发展和技术革新的相互推进；2022 年《人工智能白皮书》指出发展人工智能应考虑技术创新、工程实践和可信安全这三个维度以达成技术的持续健康发展。可见当前政策对于人工智能领域的发展和突破给予了相当程度的重视与支持。机器学习作为人工智能应用于具体场景的热点，为白皮书中指出的将知识和数据结合以达成感知智能向认知智能的转变提供了应用环境。同时，机器学习在基金等金融领域的应用亦能开拓机器学习在行业中所扮演的角色，例如生成式人工智能技术的发展可能能使机器学习不只应用于量化决策领域，更能使其进行内容的生产，从后台走向前台，进行信息输入和输出的结合，完成基金公司和用户交互的新模式。而政策的支持在给予相关科技领域足够的经费和发展动力的同时，也推动了机器学习在隐私保护、可解释性、稳定性、公平性等难点领域的发展意识，加快了技术在应用领域的完善。由此可见，政策导向对于金融科技领域达成了有效的推动，机器学习在基金应用场景的发展未来可期。

## 五、风险管理

### （一）机器学习在变量间因果识别方面面对复杂技术挑战

为了对经济社会规律进行理解、预测其运行，对变量进行因果识别是一种重要的手段。机器学习在因果识别方面可起到积极的作用，例如机器学习可通过对于数据与信息的处理对于变量间的关系进行有效的识别，因而在处理非线性的变量间复杂关系方面具有优越性等等。但在投资领域引入机器学习的新方法应当对大数据数字化背景下新问题的产生以及更复杂的因果关系分析提供具有开拓性、甚至提供传统方法难以替代的实证工具。在此过程中，风险与挑战也随之而生。

一方面，在机器学习的过程中，因果识别有被忽略的风险。在处理数据时，变量间的相关性比起因果性更易判断，而在许多时候，考察变量间是否具有相关关系对于数据预测又已显得足够充分。这增加了机器学习对于因果识别忽略的可能，与当下的主流实证方法产生了偏差。当然，这并不是说因果识别在所有情形下都是必要的或者考察变量间的相关关系难以具备现实意义，只是说明机器学习可能在保持变量处理在特定学科情形中的科学性可能会有疏漏。

另一方面，机器学习对于因果关系的识别可能产生误差或难以检验。例如，在非结构化的高维数据映射入结构化的数据的过程中，映射函数起到了至关重要的作用。在此过程中，提高识别框架的分析准确率是必要的，这对变量处理提出了高要求。但若训练集被过度贴合，机器学习的过程可能反而会吸纳本不必要的噪音和偏差，产生过拟合的问题。这使得检测出的变量间的因果性存在虚假风险。而对于因果关系的追溯和检验又有可能会由于机器学习时近似黑箱的过程而产生可解释性的缺乏，加大了检验难度。

因而，机器学习在进行预测和分析工作时，应当对因果分析环节保有足够的重视程度，即在合理的情境下应用，并充分发挥机器学习在数据化时代处理数据的优越性。同时，也可将机器学习的方法当作验证手段利用，使其应用场景更灵活，以保证基金公司在投资中使用的机器学习方法具有有效性。

## （二）机器学习的新方法受到数据来源的制约

数据作为新生产要素，是数字化的基础。在数字化背景下诞生的机器学习方法亦具有数据驱动的特征。但数据资源具有稀缺性，在该方法的普及过程中，可靠的数据资源的获取或许会因为产权等等因素而变得困难。这可能会导致两个问题：

第一，数据获取不充分的潜在风险对基金公司策略的生成和选取的有效性产生威胁。这种不充分既体现与数据的量上，亦体现于数据的质上，即数据可能并不全部公开，而其质量亦难以验证。数据的质与量有欠缺可能导致基金公司可能采取错误的投资策略。

第二，由于数据来源的重要性，行业内可能产生数据垄断的情况，而数据壁垒的产生对行业以及金融科技两者的发展与更新起到不利影响。

2022年12月2日，国务院发布了《关于构建数据基础制度更好发挥数据要素作用的意见》，对于数据产权制度、数据要素流通和交易制度等制度提出了建设意见。未来对于数据获取与流通的相关政策的落实和跟进能一定程度地缓解这一风险与问题。同时，基金公司也应对数据来源进行有效的筛选与审查，保证机器学习方法的有效性。

## （三）基金经理和投资方思维与行为习惯加大机器学习使用推广难度

受当前技术所限，机器学习在金融分析和投资领域中表现并不能在样本外的范围内应用时保持完全稳定，加之机器学习可能包含的过拟合、数据方面的技术难点、黑盒特性导致的可解释性低，以及技术开发的成本等问题较难解决现状，可能导致已成熟的基金经

理以及投资者对新技术的有效性和必要性持怀疑态度，而对传统的投资方法保持偏好，加大新技术的推广难度。

缓解此问题应在用户端适配新的沟通平台，提供相关的配套服务，使投资者对于机器学习技术及其在投资方面的优势有更充分的了解以减少怀疑和抵触情绪。

#### （四）机器学习技术发展初期对人才提出高要求

机器学习与金融领域的结合对计算机、数理和金融等多领域交叉人才提出要求。而现有的教育模式和人才储备可能难以向技术的发展领域输送足够多的优质人才。其次，新技术的变革、对人才的新要求势必导致一批人才的淘汰，因而在短期内，人才市场可能面临动荡和冲击。

因此，企业需担当部分人力资源培养的任务，对人才进行交叉学科领域的培训，推动不同领域人才的交流与合作。

## 附录

### 参考文献

- [1]周大蔚,汪福运,叶翰嘉等.基于深度学习的类别增量学习算法综述[J/OL].计算机学报:1-32[2023-04-27].
- [2]李斌,邵新月,李玥阳.机器学习驱动的基本面量化投资研究[J].中国工业经济,2019,No. 377(08):61-79.
- [3]蒋锋,张文雅.机器学习方法在经济研究中的应用[J].统计与决策,2022,38(04):43-49.
- [4]郭峰,陶旭辉.机器学习与社会科学中的因果关系:一个文献综述[J].经济学(季刊),2023,23(01):1-17.
- [5]纪守领,杜天宇,李进锋等.机器学习模型安全与隐私研究综述[J].软件学报,2021,32(01):41-67.
- [6]彭方平,王茹婷,廖敬贤.银行竞争、企业债务与“脱虚向实”——基于双重机器学习方法[J].经济理论与经济管理,2023,43(03):82-97.
- [7]倪宣明,郑田田,赵慧敏.军民融合能降低高新技术企业权益融资成本吗?——基于双重机器学习的实证研究[J/OL].系统工程理论与实践:1-22[2023-04-27].
- [8]中国证券投资基金业协会.(2021).中国证券投资基金业年报2020.[R]北京:中国财富出版社。
- [9]中国证券投资基金业协会.(2020).中国证券投资基金业年报2019.[R]北京:中国财富出版社。
- [10]Grobys, K., Kolari, J. W., & Niang, J. (2022). Man versus machine: on artificial intelligence and hedge funds performance[J]. Applied Economics, 54(40), 4632-4646.
- [11]吕晶晶.公募量化基金的布局与分化[J].金融博览(财富),2022(05):47-49.
- [12]周方召,石祥翔,贺志芳等.量化投资基金和股票市场稳定性[J].金融经济研究,2023,38(02):81-96.