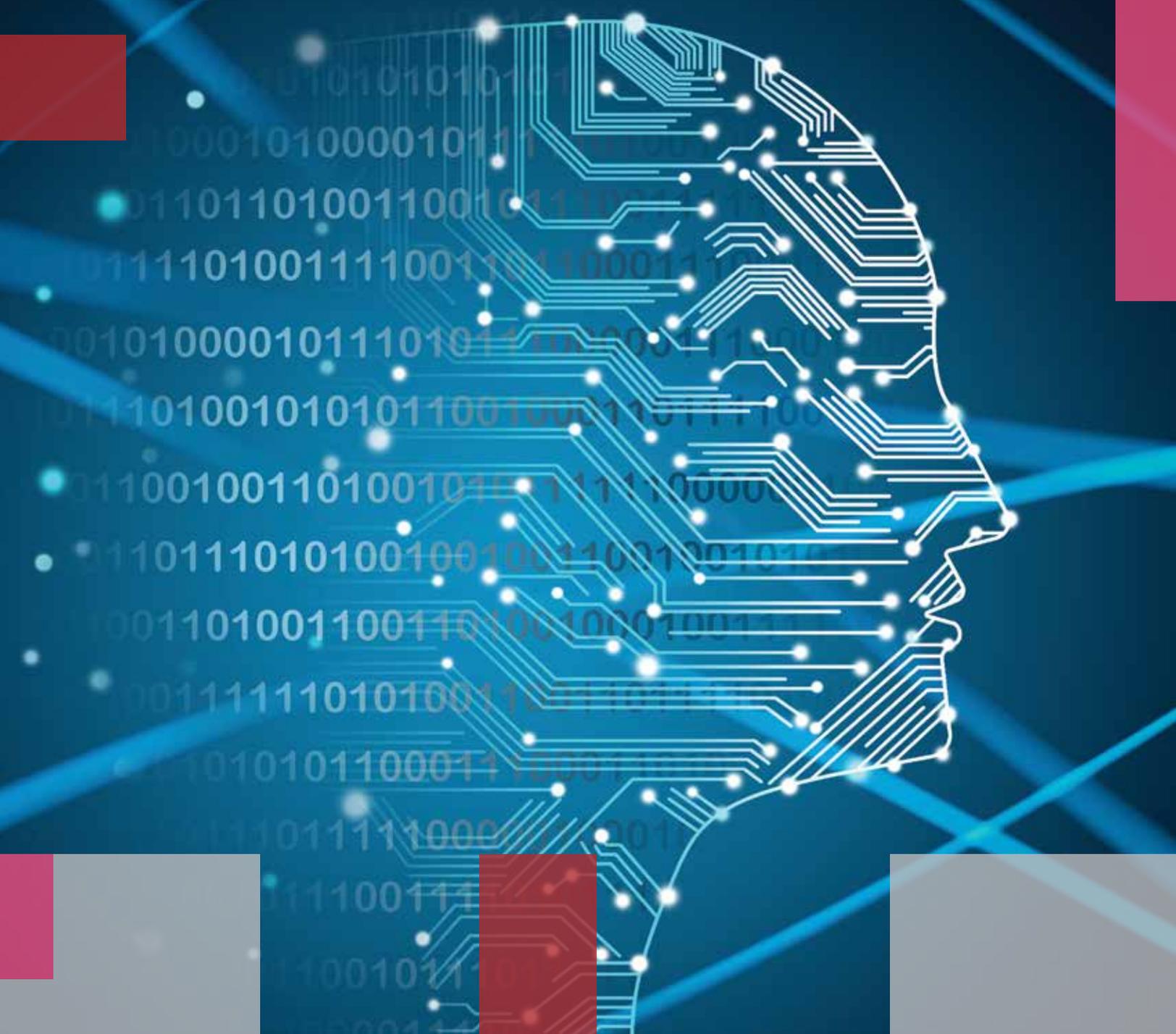




CFA Institute

投资管理界的 AI 先行者

检视人工智能和大数据技术
在投资领域的趋势和用例



© 2020 CFA Institute. All rights reserved.

本出版物版权归属 CFA Institute。未经版权方事先书面许可，不得以任何形式或方式（包括但不限于电子或机械复印、录制及信息存储与检索系统）翻印或转载。本出版物旨在提供与所涵盖主题有关的准确且权威的信息。本出版物的传播建立在读者了解出版方不提供法律、会计或其他专业服务的基础上。因此，如需专业咨询或其他专家协助，请寻求相关合格专业人士的服务。

CFA[®]、Chartered Financial Analyst[®]、Investment Foundations[™] 及 GIPS[®] 均为 CFA Institute 注册商标。查看 CFA Institute 所属商标完整列表与使用指南请访问 [HYPERLINK www.cfainstitute.org](http://www.cfainstitute.org)。

免责声明

本报告旨在提供与所涵盖主题相关的准确且权威的信息。本报告的传播建立在读者了解出版方不提供法律、会计或其他专业服务的基础上。此外，出版方不提供任何投资建议，亦不为任何方法论背书。本报告中包含的案例并不涉及对案例中提及人员、产品或企业的背书；投资产品的回报不可预见。如需专业咨询或其他专家协助，请寻求相关合格专业人士的服务。

ISBN: 978-1-942713-78-4

投资管理界的 AI 先行者

目录

摘要	4
前言	5
人工智能和大数据在投资领域的应用：挑战与机会	7
人工智能和大数据在投资领域的应用：新前沿	7
投资行业在应用人工智能和大数据方面的发展	8
应用人工智能和大数据的挑战：金融科技金字塔	14
化为现实：T 型团队	16
展望与警示	17
案例研究	18
1. 通过机器学习增强交易策略和执行：英仕曼 AHL	18
2. 利用机器学习生成量化分析模型的信号：纽约人寿投资公司	20
3. 通过深度学习完善股票成交量预测：道富集团	22
4. 在卖方研究中利用人工智能 / 另类数据分析：高盛	24
5. 借助人工智能和大数据剖析财报电话会议：美国世纪投资公司	26
6. 人工智能和大数据助力债务投资组合管理：中国人寿资产管理公司与 中证信用增进股份有限公司	28
7. 运用人工智能和大数据技术受理和处理保险索赔以及评估企业风险：平安集团	30
8. 情绪分析：彭博	32
9. 建设数据科学团队：施罗德	34
10. 专题焦点：通过机器学习增强 MPT 的有效边界	37
11. 专题焦点：使用智能搜索采集和处理信息	40
致谢	42

摘要

机器人是否会取代人类投资经理？可以说，投资行业正处在最重大的技术变革关口，我们希望能够了解投资管理行业采用人工智能（AI）和大数据技术的现状，并能够举一反三，找出这些技术可能的用武之地和发挥作用的方式。

我们发现，目前在投资流程中利用人工智能和大数据技术的投资从业人员还相对较少。为了给希望利用最新前沿技术的投资机构和个人提供一个标杆，我们与全球范围内目前在采用这些技术的部分机构进行了深入交流；它们都是投资管理界的 AI 先行者。

他山之石，可以攻玉。本报告中呈现的案例具有极高的借鉴价值。这些案例不仅凸显了人工智能将为我们带来的机会及局限所在，也印证了人类的判断力在投资过程中依然具有的重要作用。

我们将此归结为“AI + HI”模式的力量：人工智能可以增强人类智慧，从而帮助投资从业人员提高业绩，从日常重复性的工作中解放出来，借助机器和人类的集体智慧，作出更明智的决策。

未来的成功将属于那些在战略规划上将人工智能和大数据技术纳入投资流程的投资机构，也属于那些理解并充分利用这些新技术带来的机会的投资从业人员。

未来已来。

主要结论

- 使用人工智能和大数据技术的决策，应当与传统技术的表现进行对标。企业应确定，与运用人工智能和大数据的额外成本和复杂度相比，获取潜在的额外阿尔法收益是否值得。
- 机器的智能程度取决于它所学习的数据。用于训练的数据越综合全面，机器对新事件的处理将会越具有概括性，从而减少过拟合等常见的陷阱。
- 机器学习技术更适合系统性的策略（包括基于规则的策略、量化策略），全权委托（主动管理型）投资经理通常更多使用非结构化数据和另类数据。
- 精细的行业特定数据集更适合基本面分析师或追求阿尔法收益的基金经理，而不适合采用系统性策略的投资经理。
- 有效使用此类数据集，可以为“四面楚歌”的主动管理行业带来极大的机会。
- 人工智能和大数据不是万能钥匙；不能解决所有的投资问题。例如，只有少部分的大数据能够生成有意义的信号；从噪音中可靠地提取信号是有难度的。

要点

- 我们识别了人工智能在投资管理领域的三个用途：(1) 借助自然语言处理（NLP）、计算机视觉和语音识别技术高效地处理文本、图像和音频数据；(2) 借助机器学习（ML）技术（包括深度学习技术）提高投资流程中使用算法的有效性；(3) 借助人工智能技术来处理大数据（包括另类数据和非结构化数据）从而得出投资见解。
- CFA Institute 的调查显示，目前在投资流程中使用人工智能 / 大数据技术的投资从业人员仍然相对较少。大多数基金经理依然依赖 Excel 表格和电脑版市场数据工具；在过去 12 个月里，仅 10% 的受访基金经理使用过人工智能 / 机器学习技术。
- 我们发现，对于人工智能和大数据在投资流程中的成功应用，存在以下五个主要障碍：成本、人才、技术、领导层愿景和时间。投资机构需要大力攻克这五个障碍，才能到达金融科技金字塔的顶端。
- 强大的金融科技是金融（金融机构）与科技（科技企业）合作的产物。成功的企业将以 T 型团队为中心，这样的团队是在整个投资策略或流程中综合投资专业实践、创新和技术应用。

前言

在本报告中，我们希望通过分析具体的用例，识别人工智能（AI）和大数据在投资领域中具有深刻影响力的应用场景，以及运用这些技术的最佳实践。为此，我们采访了来自不同国家和地区以及不同领域的投资从业人员，采访的时间主要是在 2019 年 4 月和 5 月。

我们为什么要推出本报告？

正如我们在以前的 CFA Institute 研究项目（最近的为《2018 亚太金融科技概览》¹）中所指出的，金融科技的早期阶段（例如网贷、移动支付和智能投顾）对金融服务业的补充性大于颠覆性。但金融科技的“ABCD”技术，即人工智能（AI）、区块链（Blockchain）、云计算（Cloud Computing）和大数据（Big Data）等新技术可能会令金融服务行业发生重大转变。

本报告重点关注的是人工智能和大数据，这是投资管理行业科技发展的前沿。区块链和云计算与技术基础设施的关系比在应用层面的更紧密一些，不在本次调研范围内。

自《2018 亚太金融科技概览》报告发布以来，通过与全世界多家主要金融机构和研究机构的技术高管和资深研究人员交流，我们感觉人工智能和大数据在金融和投资领域的应用已经走下“期望火热的巅峰”，进入“幻想破灭的低谷”。²

为帮助那些目前在探索不同的人工智能和大数据应用的投资从业人员，以及计划开始探索如何应用这些技术但还不知道从哪里入手或如何入手的投资从业人员，我们找出了一些人工智能和大数据领域的代表性案例。

本次调研的目的，是为了明确相关的用例，在人工智能技术重塑投资管理行业发展的形势下，让投资从业人员和投资机构可以采取恰当的行动，应对不断变化的行业格局，为投资成功做好准备。

本报告包含哪些内容？

在《2018 亚太金融科技概览》中，我们详细地讨论了金融科技的 ABCD 技术（尤其是人工智能）可能对金融服务业的转型带来哪些变化。在《未来投资专业人士》³中，我们讨论了面对技术变革的大环境，未来投资团队的关键角色和技能。在本报告中，我们将在这两条思路的基础上进一步延伸，介绍一些投资机构在将人工智能和大数据技术纳入投资流程方面已经取得的成绩，以及这些变革是如何成为现实的。对于每个案例，我们重点关注三个领域：

1. 投资流程发生了哪些变化？
2. 哪些人工智能和大数据技术推动了这些变化的实现？
3. 团队中的关键角色（以及要求的相关技能）是什么？团队如何协作以实现改变？

我们相信，实现以上目的的最佳方式，是从全球信誉良好的投资机构中找到在实践中应用人工智能和大数据技术的切实案例。在选择本报告中的案例时，我们遵循如下标准：

- 地域性——案例来源机构的业务应覆盖三大区域（美洲地区、亚太地区和欧洲地区）
- 专业性——涵盖投资业务的四个主要领域：股票、债务、资产配置和对冲基金（另类投资）
- 实践应用性——仅选择实际进行应用的案例，不考虑概念验证

这些案例还包括全权委托和系统性策略。我们的调研结果证实，机器学习技术更适合系统性的策略，而大数据更多被全权委托投资经理使用。

1 曹实，《2018 亚太金融科技概览》（弗吉尼亚州夏洛茨维尔：CFA Institute，2018 年）。

2 Gartner, Gartner Hype Cycle. www.gartner.com/en/research/methodologies/gartner-hype-cycle.

3 CFA Institute, 《未来投资专业人士》（弗吉尼亚州夏洛茨维尔：CFA Institute，2019 年）。<https://futureprofessional.cfainstitute.org/>.

此外我们还包含了两个不属于“核心”投资管理领域的案例——分别为高盛（专注于卖方研究业务）和平安集团（一家以保险业务起家的中国金融集团）。随着 MiFID II 等行业变革，卖方研究的价值正受到越来越多的质疑。卖方分析师如何证明其研究的力量？高盛的案例可能为投资经理提供一些借鉴。平安集团案例的强大之处不仅在于其人工智能团队的规模和高速增长，也在于人工智能和大数据在其核心业务中的广泛应用。

另外还有两个专题焦点案例，更深入地分析了人工智能和大数据在投资领域的具体应用。Marcos López de Prado 的案例讨论了如何借助机器学习技术帮助提高 MPT 投资组合的业绩和稳定性。杨永智分享的案例介绍了如何使用自然语言处理（NLP）技术来采集和处理海量信息。

总之，这些案例涵盖了从投资机构小型团队到金融集团大规模技术团队的运作。通过这种多样性的展示和分析，我们希望能够让各种规模的企业都能从我们的发现中受益。

我们选择的案例注重以下方面的多样性

- 地域
- 专业
- 实践应用
- 机构 / 团队规模
- 人工智能 / 大数据成熟度

人工智能和大数据在投资领域的应用：挑战与机会

人工智能和大数据在投资领域的应用：新前沿

用许多行业领军人物的话说，人工智能就是新电能。《经济学人》杂志曾宣称认同“数据就是新时代的石油”。⁴在这一部分，我们讨论了一些机构如何在投资管理中发挥前沿技术的力量。

人工智能：NLP、计算机视觉和语音识别。

近年来，研究人员在构建“能看、能听、能理解”的终极机器方面取得了长足的进步。⁵在自然语言处理（NLP）、计算机视觉和语音识别程序方面，人工智能被用于从各类公共来源以及内部 / 供应商数据库提取文字、音频和图像。例子包括分析师电话会议转录，以及从发行人申报资料中提取数据以用于估值模型。在大多数情况下，程序会将传统上由分析师手动进行的重复工作自动化。我们预计，这些类型的应用将在行业中越来越普遍；通过梳理多个数据来源并将它们整合到一个平台，扩大了投资从业人员的范围，提高了他们的效率。

这些程序还节省了在手动工作上花费的时间，从而提升了人的能力。初级分析师的很大一部分时间都用在寻找和录入信息上。这些重复的例行工作很可能最先被人工智能程序接管，因为后者在此类工作上拥有先天的优势。

人工智能：机器学习与深度学习

更复杂的程序会进一步处理来自不同渠道的信息，从而向投资决策流程输出相关信号。这往往需要机器学习和深度学习等复杂的人工智能技术。

机器学习是一个通用名词，指可让机器在没有明确编程指令的情况下发现模式的计算方法和算法。⁶机器学习程序会自行处理如何解释输入并预测输出。⁷深度学习是机器学习的一种，它基于神经网络（一种以人脑为模型的学习）。

深度学习算法常常用于改进 NLP、计算机视觉和语音识别程序的结果。它们还可帮助从海量的数据中提取有用信息。例如，这些算法可以从电话会议转录文稿中推导出某些关键词，或者从社交媒体等非结构化数据中识别情绪。然后将这些信息化为交易信号或预警，以供人类分析师和基金经理处理。

机器学习和深度学习程序在量化（系统性）投资经理中也很受欢迎，他们发现运用这些技术对提高量化投资流程的效果极有帮助。本报告中的多个案例也证明了这一点。

传统统计学和计量经济学基于最早于数个世纪前开发的技术，它们在金融领域的应用也经常会涉及到线性回归模型。这些线性模型在许多情况下都非常有效。由于机器学习技术能够处理经常会在金融领域出现的上下文和非线性关系，使用这些技术或可以更好地把握真实世界中至少一部分复杂性。例如，存在多重共线性（解释变量之间有相关性）时，机器学习技术比线性回归更有效。⁸在这种情况下，机器学习和深度学习技术为投资经理提供了额外的工具包，让他们可以具备一定的竞争优势。⁹

4 参见 www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data。

5 参见曹实，“Artificial Intelligence, Machine Learning, and Deep Learning: A Primer”，Enterprising Investor 主题博客（2018年2月13日）：<https://blogs.cfainstitute.org/investor/2018/02/13/artificial-intelligence-machine-learning-and-deep-learning-in-investment-management-a-primer/>。

6 K.C.Rasekhschaffe 和 R.C.Jones，“Machine Learning for Stock Selection”，《Financial Analysts Journal》第 75 卷，第 3 期（2019 年第三季度刊）。

7 曹实，“Artificial Intelligence, Machine Learning, and Deep Learning”（2018 年）。

8 参见 Rasekhschaffe 和 Jones（2019 年）。

9 有关机器学习的延伸阅读建议，包括机器学习算法及其应用的类型，参见 CFA Institute Refresher Reading 系列有关“机器学习”（2020 年版教材），CFA 持证人员可通过链接 www.cfainstitute.org/en/membership/professional-development/refresher-readings/2020/multiple-regression-machine-learning 获取。

大数据：另类数据与非结构化数据

数据科学家将大数据定义为四个“V”：大量 (Volume)、多样 (Variety)、真实 (Veracity) 和高速 (Velocity)。¹⁰ 在投资圈经常使用的术语则是“另类数据”或“非结构化数据”。

另类数据是指来自当前尚未使用或还不属于主流的渠道的数据。与数字化后在关系数据库中存储的结构化数据相反，非结构化数据是指通常以文本、图像或语音格式存在，并且无法方便处理的数据。另类数据与非结构化数据之间存在关系，但并不完全相同。另类数据在初次发现时往往是非结构化的，非结构化数据通常不会被主流投资者使用，因此是另类的。

当今投资界经常使用的另类数据 / 非结构化数据例如卫星图像、财报电话会议录音和转录文稿、社交媒体发布内容、消费者信用卡和借记卡数据以及电子商务交易等。

寻找能够创造阿尔法收益的新数据来源，已经成为一些分析师和投资经理的新一轮“军备竞赛”，与过去投资经理努力在公开市场中千方百计地寻找“未垦之地”类似。在某种程度上，从大数据中提取信号就是分析师们一直在做的工作的延伸——例如走访门店检查客流量。现在，一些分析师可以使用卫星图像或停车场的传感器信息来进行这样的推算。新技术往往可以提高效率；通过卫星图形或传感器数据，分析师可以用更少的时间覆盖更多的门店。

另类数据往往是小众的，也更受基本面型投资经理欢迎，他们将数据作为投资决策流程中的一个信息输入。本报告中的一些案例提供了如何使用另类数据和非结构化数据的真实操作。

人工智能 / 大数据在投资领域的应用类型

- **人工智能：NLP、计算机视觉和语音识别。**用于处理文本、图像和音频数据。
- **人工智能：机器学习。**用于提高投资流程中所用算法的效果。
- **大数据：另类数据与非结构化数据。**用于处理另类数据与非结构化数据，从而获得投资见解。

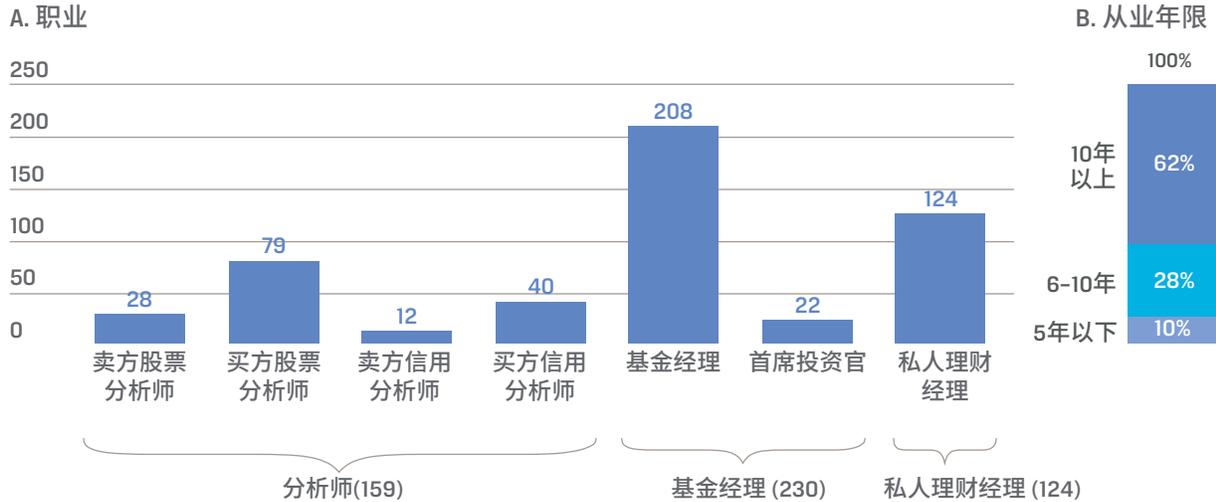
投资行业在应用人工智能和大数据方面的发展

CFA Institute 开展了一次实践分析调查，以了解在分析师、基金经理和私人理财经理的工作流程中采用不同技术的情况。这一部分提供了一些相关的发现，以说明投资行业在人工智能技术应用方面的格局，为后面的具体案例研究确立背景。

调查问卷在 2019 年 3 月发送至随机抽样的 CFA 持证人，总计有 734 名受访者（其中 52% 来自美洲地区，18% 来自亚太地区，30% 来自欧非中东地区）。受访者的职业包括股票和信用分析师、基金经理、首席投资官以及私人理财经理，如图 1 中 A 部分所示。受访者的从业经验如 B 部分所示。

¹⁰ 参见 www.ibmbigdatahub.com/infographic/four-vs-big-data。

图 1：受访者统计情况



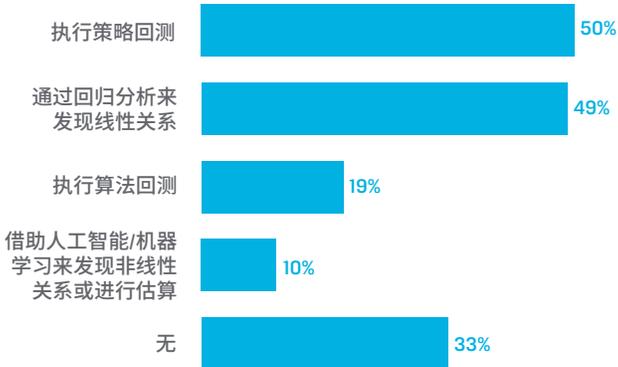
注：调查参与者总数 (N = 734)。

调查结果显示，目前很少有投资从业人员使用通常在机器学习技术中应用的程序，例如 Python、R 和 MATLAB 等编码语言。大多数基金经理在投资策略和流程上依然依赖 Excel 表格 (占受访基金经理的 95%) 和电脑版市场数据工具 (占受访基金经理的四分之三)。

此外，如图 2 所示，在过去 12 个月里，仅 10% 的受访基金经理使用过人工智能 / 机器学习技术，在投资策略和流程中使用线性回归的受访者数量，几乎是使用人工智能 / 机器学习技术的受访者的五倍。

图 2：投资策略和流程中使用的统计技术

基金经理：过去12个月里，您在投资策略和流程方面使用了以下哪些方法？



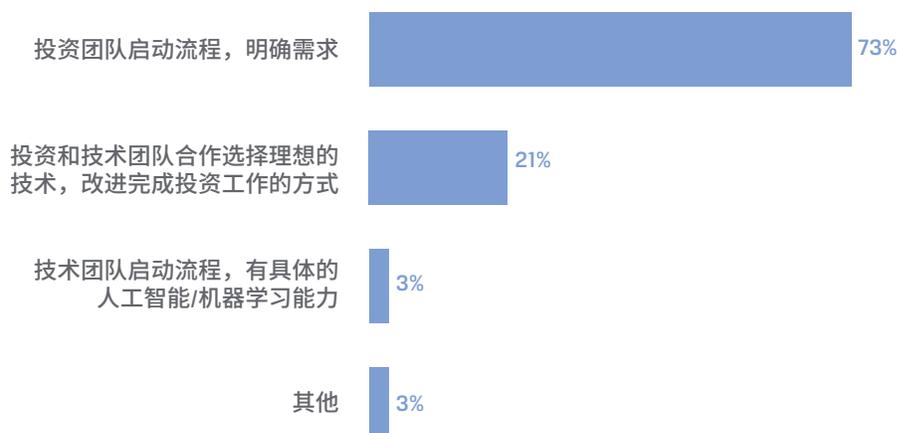
注：调查参与者总数 (N = 230)。

参与调查的受访基金经理中，仅 10% 在过去 12 个月内使用人工智能 / 机器学习技术来改进他们的投资流程。

在公司组织层面，投资团队和技术团队的协作水平 可能需要进一步的整合才能实现流程效率的提升。依然较低,如图 3 所示。这说明随着这些技术的普及,

图 3：公司对投资策略和流程的职责

基金经理：以下哪项最能反映所在组织的投资策略和流程？



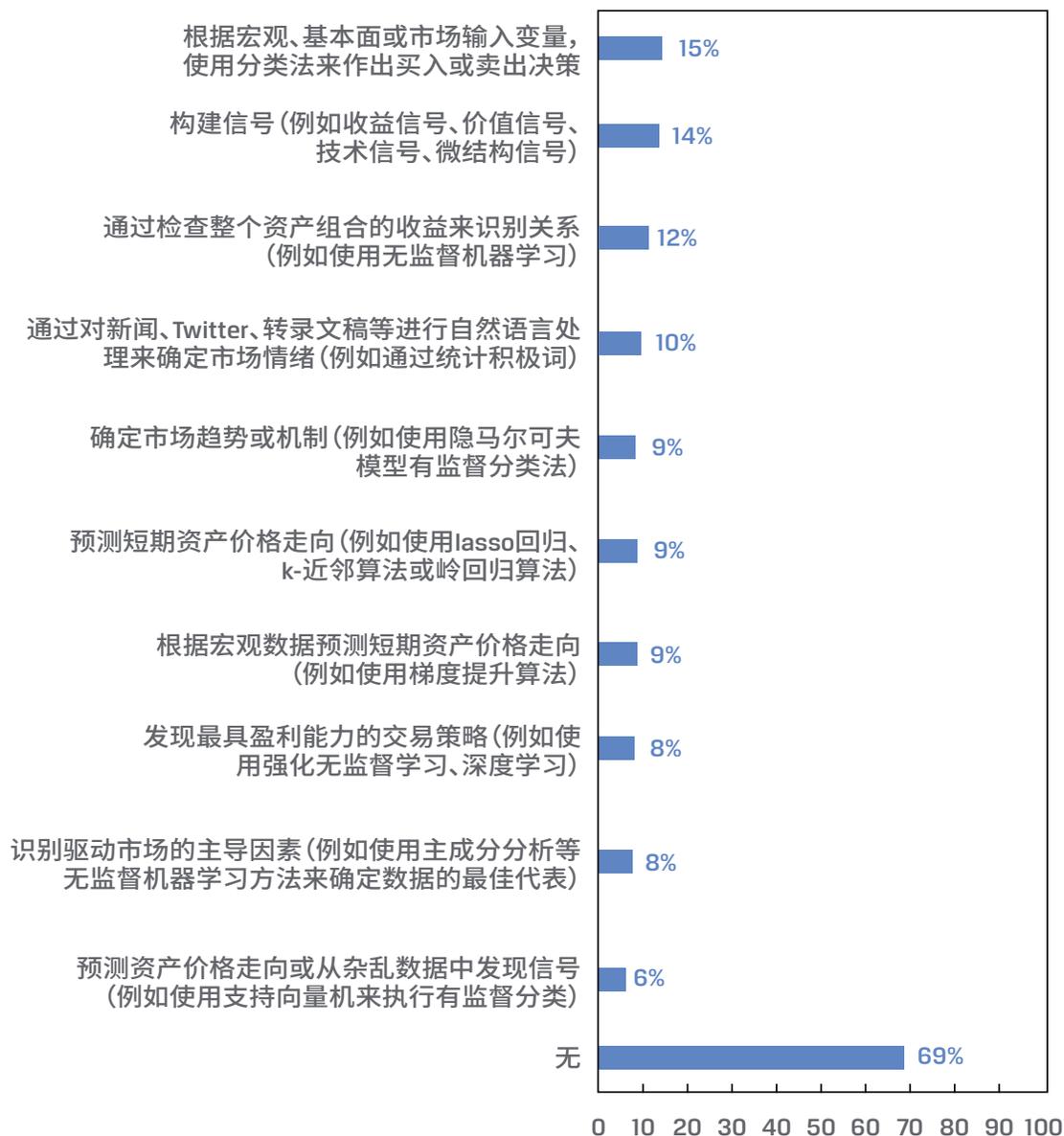
注：调查参与者总数 (N = 230)。

调查显示，人工智能 / 机器学习技术在交易策略中的普及程度也较低。如图 4 所示，69% 的受访基金经理表示，在过去 12 个月内未使用任何人工智能 / 机器学习技术来创建交易算法。

正在使用这些技术的从业人员则表示，他们的使用范围十分广泛，包括根据若干数据变量作出买入或卖出决定 (15%)、构建信号 (14%) 以及根据 NLP 确定情绪 (10%) 等等。

图 4：创建交易算法时对人工智能 / 机器学习技术的使用

基金经理：在过去12个月里，您使用了以下哪种人工智能/机器学习技术来创建交易算法？



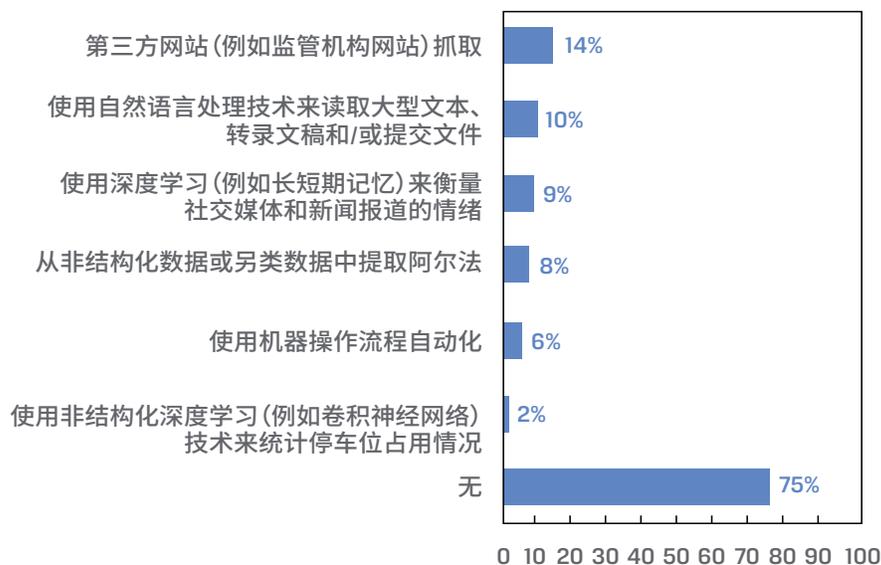
注：调查参与者总数 (N = 230)。

图 5 中 A 部分显示类似的结果，四分之三的受访分析师没有将人工智能 / 机器学习技术用于行业和公司分析。在使用这些技术的受访者中，最受欢迎的两个技术分别是抓取第三方网站 (14% 的受访者表示使用) 和使用 NLP (10% 的受访者表示使用)。而有 40% 的受访者将线性回归用于行业和公司分析 (未显示)。

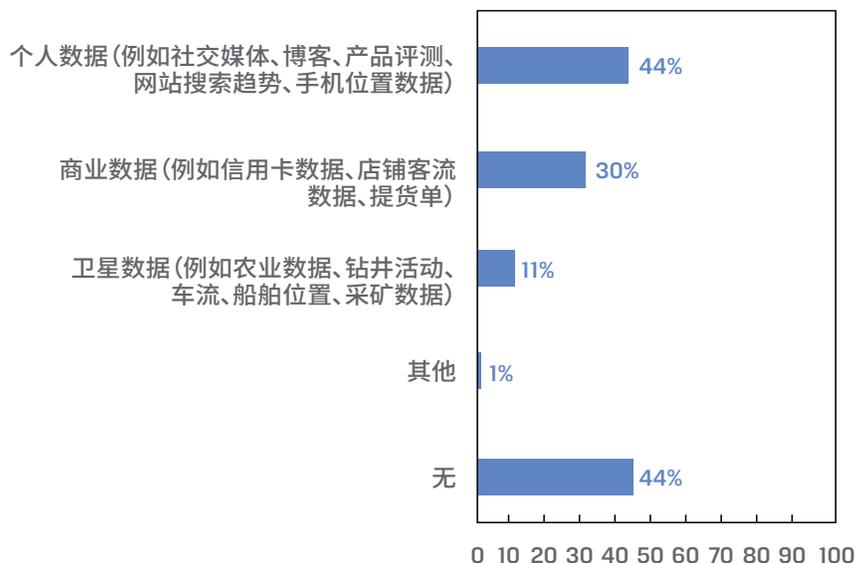
在投资从业人员中，将非结构化数据和另类数据用于行业和公司分析，比使用人工智能 / 机器学习技术更流行。如图 5 中 B 部分所示，44% 的受访分析师表示在过去 12 个月内使用个人数据，例如社交媒体、产品评测和网站搜索趋势等，而仅 11% 的受访者曾经使用卫星图像。但这些结果的一个弊端是，它们无法让我们推导这些数据来源于行业和公司分析中的使用频率和密集度。大量的从业人员 (44%) 表示未曾使用这些数据。

图 5：行业和公司分析中对人工智能 / 机器学习技术与非结构化数据 / 另类数据的使用

分析师：在过去12个月里，您将以下哪些人工智能/机器学习使用情形用于行业和公司分析？



分析师：在过去12个月里，您将哪些类型的非结构化数据和/或另类数据用于行业和公司分析？



注：调查参与者总数 (N = 159)。

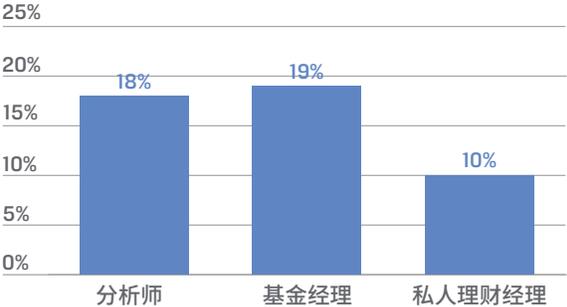
总之，这些结果说明投资行业在人工智能及相关技术的采用方面仍处于极为早期的阶段，目前很少有从业人员在日常投资流程中使用人工智能 / 大数据技术。

总体上，由于目前人工智能和大数据技术的利用率低下，加上参与这些领域培训的从业人员数量庞大，行业似乎将在未来几年迎来明显的增长。

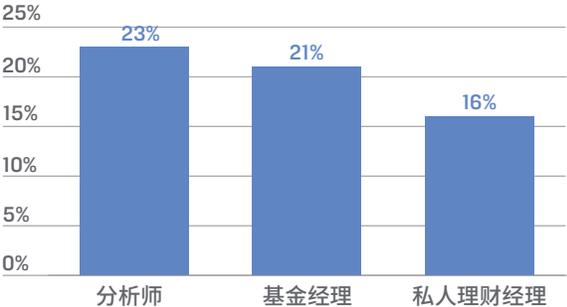
但大约有五分之一的分析师和基金经理表示参加了人工智能 / 大数据培训，如图 6 所示。

图 6：人工智能 / 机器学习与数据分析培训

A. 过去12个月内完成的人工智能/机器学习培训

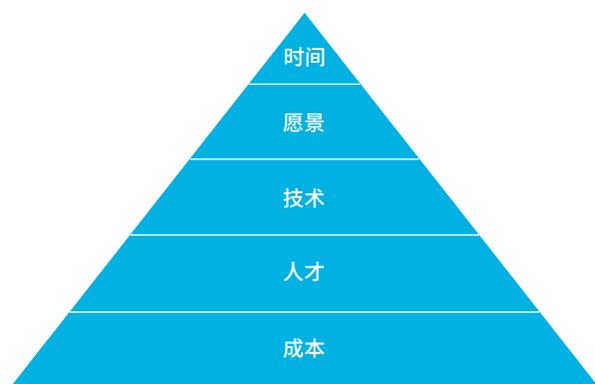


B. 过去12个月内完成的数据分析培训



应用人工智能和大数据的挑战： 金融科技金字塔

图 7：金融科技金字塔



那么，哪些因素导致投资从业机构和投资机构没有充分发挥人工智能和大数据的威力？我们认为有五个主要障碍，按从易到难的顺序列举如下。

障碍 1：成本

金融机构对巨额的 IT 预算并不陌生，但开展人工智能和大数据应用可能会涉及极大的前期成本以及持续维护成本。

这一高成本至少可部分归结于支持这些技术应用且一直受到行业关注的新数据集。¹¹ 发现、清理和整理这些数据集绝非易事，为此一位著名的经济学家认为，在人工智能和大数据时代，小型机构会越来越难以参与竞争。¹²

障碍 2：人才

在人工智能时代，拥有基本编程和统计教育背景的大学毕业生非常受用人单位欢迎，更别提那些拥有人工智能或相关领域高级学位的毕业生。而这仅仅是问题的一个方面。

在采用并大力投资人工智能的顶尖科技公司工作，将拥有切实的优势，例如谷歌、微软、百度和阿里巴巴等等。人工智能领域许多最新和最重大的发展都发生在这些公司，参与这些项目的少数员工已经成为“稀缺物种”，他们可以接触到目前全世界顶尖大学还没有教授的知识和技能。然而，实际进入投资行业工作的顶尖人工智能人才似乎极为稀少，这使问题进一步复杂化。也许 AlphaGo 和无人驾驶汽车天生更能令人工智能科学家感到兴奋。不论哪种途径，获取人才的难度看起来至少要比管理成本高一级。

障碍 3：技术

人工智能革命方兴未艾，技术仍在快速发展。由于后来者居上的可能性极大，这一潜在风险给投资人工智能应用的机构带来极大的挑战。跟上最新发展趋势是大多数投资从业机构和机构面临的一个真实挑战，只有极少数能做到。拥有雄厚的预算和顶尖的人才才是保持领先的前提条件。

同样，在另类数据领域，对新数据来源的利用也处于较新的时期。许多新的数据供应商进入了市场，从雪崩般的数据中提取有用信号仍然极具挑战。

障碍 4：愿景

未来数十年，人工智能和大数据技术的进步可能会给投资行业带来天翻地覆的变化。这些技术变革必须由公司最高层掌舵，以在整个企业全面普及，同时高效地部署资源。

正如《Investment Firm of the Future》报告中所指出，迄今为止，投资机构的 IT 部署一直很被动，只是尝试利用技术手段解决历史遗留问题，提升效率。¹³ 投资机构需要积极主动地发展技能和采购系统，以保持竞争优势。战略愿景、领导层的承诺以及对 IT 部署的集体责任感，是未来企业成功的关键。

¹¹ 例如，参见 www.cnn.com/2017/11/28/making-millions-from-the-data-hidden-in-plain-sight.html。

¹² www.nytimes.com/2018/01/12/business/ai-investing-humans-dominating.html。

¹³ CFA Institute, 《Investment Firm of the Future》(弗吉尼亚州夏洛茨维尔：CFA Institute, 2018 年)：第 10-11 页。

障碍 5：时间

任何进步，不论多小，除其他努力之外，往往都需要极大的时间投入。这是身处发展前沿时自然会遇到的真实情况。

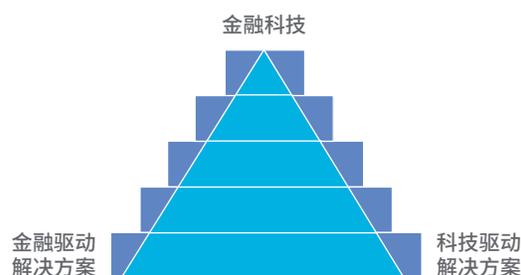
每个企业都希望做第一家通过探索和努力发现有用信息的组织，但寻求提高阿尔法收益的途径并将新方法融入现有的投资流程都非一夕之功。即使在最成熟的市场，在已经部署最先进技术多年的企业中，大多数大数据项目仍然需要极大的时间和投入来准备数据，使其满足项目的需求。耐心和坚持是必要条件，但尽管如此，许多项目也不会成功。时间依然是最难克服的挑战之一，成功不会在朝夕之间出现。

当前进展与登顶金字塔的道路

投资机构需要大力克服这五个障碍（即使用最新的人工智能和大数据技术来解决核心的投资问题），才能达到金字塔的顶端，也就是“金融”与“科技”相会之处。但登顶之路非一家之力能够成就，需要通力协作；每个障碍的克服都需要从金融和科技两个维度考虑问题。

图 8 展示了这个概念。在“金融”一角，投资解决方案往往由具有金融背景的量化分析师推动。总体而言，此类解决方案依赖现有的数据集，不是特别依赖另类数据，节省了识别相关性（将信号与噪音区别开来）以及测试和清理数据的时间与精力。此外，他们可能不会受益于来自另类数据以及 NLP、计算机视觉和语音识别等最新技术突破的新信息。

图 8：金融与科技相会之处



此外，一些主动式基金投资经理正在试验他们遇到的新数据来源。常被忽视的是系统性地利用新技术来收集和将输入到投资流程的新信息，来建立竞争优势的整体战略。

而在“科技”一角，解决方案一般由来自投资界以外的技术专家推动。架构师和他们的团队往往十分精通最新的人工智能和大数据技术，可以创造利用最新技术的“神奇魔法”。但这些成果往往不会在意具体的业务目标或最终用户特征，无法方便地纳入成熟投资机构的投资流程中。

将人工智能和大数据引入投资领域，可能是投资从业人员在其职业生涯中将遇到的投资流程中最大的变化。金融科技金字塔说明了这种复杂度，攀登者可能需要经过许多次的迭代，才能顺利登顶。重要的经验是需要采取协作式的方法，做好一步一步攀登金字塔的准备，这个过程中没有捷径。

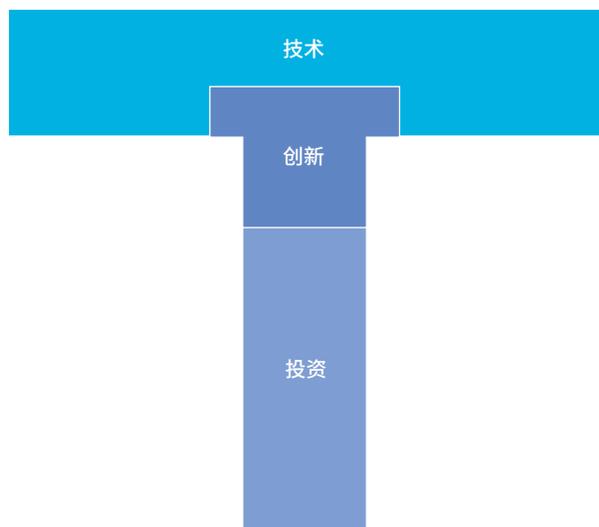
化为现实：T 型团队

金融科技金字塔凸显了在投资领域应用人工智能和大数据的战略紧迫性。我们在这一部分提出的 T 型团队概念，提供了实现这一目标的操作和组织方法。

在《未来投资专业人士》¹⁴ 中，我们讨论了投资从业人员个人取得 T 型技能的重要性越来越高。T 型专业人士不仅掌握本职领域特定的专业知识，同时拥有更广的专业连接、理解和组织视角。此外，T 型团队拥有广泛且深刻的集体智慧，享受协作式的文化和认知多样性。

对于投资领域的人工智能和大数据，我们可以运用 T 型团队的概念。将通过跨投资策略或流程的投资专业能力和技术应用取得的技能和集体智慧结合，正是 T 型团队在此领域的一个例子。在此我们还强调 T 型团队的第三个方面，也就是“创新者”在连接投资和技术团队方面的作用，要求这些专业人员尤其具有良好的 T 型技能。图 9 中大 T 形中的小 T 形就代表这一方面。

图 9：T 型团队



在此框架下，投资职能中的职位角色与当今行业的情况没有显著差异，但技术职能则可能不是如此。未来投资团队中的技术职能可能需要具备与今天不同的技能组合。除计算机工程师外，数据科学家将变得尤其重要。

第三个职能（创新）将至关重要，因为它的主要作用是促进投资职能和技术职能之间的协作，这是目前投资行业比较薄弱的环节。创新者的头衔可能是研究员、策略师、产品经理或业务开发者。在不同的机构，承担这一职能的从业人员往往分属不同的部门，对此职能的不重视可见一斑。不论是来自投资部门还是技术部门，必须清晰地了解对接部门的业务，才能成为高效的协作者。例如，策略师和产品经理承担这一职能，可能是因为他们比来自投资和技术职能的专业人员更了解大局。

正如《FinTech 与金融服务业的未来》一文中所指，强大金融科技是金融（机构）与科技（公司）强强联合的产物。¹⁵ 不论是从战略的角度还是从操作和组织角度，道理都是一样的。

在协作的早期阶段，T 型团队一般规模较小，往往仅以非正式的项目特定团队形式出现。随着运作的逐渐成熟，T 型团队成为组织结构中更常见、更持久的职能。当前问题的复杂度要求组织作出承诺，通过组织支持的 T 型团队数量（和有效性）最能看出。

¹⁴ CFA Institute, 《未来投资专业人士》（弗吉尼亚州夏洛茨维尔：CFA Institute, 2019 年）。

¹⁵ 曹实, 《FinTech 与金融服务业的未来》, 《走近互联网金融》（弗吉尼亚州夏洛茨维尔：CFA Institute, 2017 年）。中文版见 www.i.hkej.com/dailynews/investment/article/1313726/。

展望与警示

投资管理中的人类与机器视角

人工智能和机器人是否会智能到足以取代我们？恐惧在人类心理上挥之不去，弥漫在目前有关金融科技讨论的讨论中。¹⁶

我们坚信“AI + HI”（人工智能 + 人类智能）模式的强大力量——即大多数任务仍将最适合人工智能和人类智能结合处理，两者的集体力量大于任何单独一方。¹⁷ 实现这一模式的路径从日常的基本任务开始，例如从文本和图像中提取信息、生成报告、填写电子表格模型等，人工智能在能够高速处理的信息广度上比人类有优势。分析师就可以从中解放出来，负责需要更多经验和判断的更高价值的任务。

这不是人类与机器之间的竞赛。最终的竞争在于“AI + HI”团队之间，有效利用和综合两者的强队最终将打败弱队。未来成功的投资团队将通过认知多样性（人工和人类）和 T 型技能来发挥集体智慧。

在投资管理领域应用人工智能和大数据的视角

尽管人工智能和大数据将在投资行业发挥重要作用，但它们不是万能钥匙。在一些情况下，额外的信息（大数据）可以增加阿尔法收益，而在另外一些情况下，增强的算法（机器学习）可能会发现以前未发现的模式。同样，人工智能和大数据肯定不能提供投资者需要或希望的所有答案。

例如，机器学习技术面临的一个挑战是它们在测试环境中效果更好（即基于训练数据集），可能不会有效处理真实世界的新情况。

这是过拟合的问题，也就是算法在样本中表现良好而在样本之外表现不佳。人工智能可能适合 AlphaGo，因为所有规则都是确定的。而不断变化的投资世界则带来更多难题。此外，至少有些商业化的机器学习程序其实更像黑盒子；用户无法接触机器学习操作背后的逻辑。这导致程序捕获的一些特征与模型试图预测的变量之间没有因果关系。

随着技术的进步和理解的加深，这些挑战可能会被克服，但在目前，我们应在开始探索未知世界的时候将人工智能和大数据的力量考虑进来。

展望

根据我们的研究，包括对投资和科技领域学术界以及从业人员的采访与交流，我们可以得出如下结论

1. 人工智能和大数据有可能给投资管理行业带来当前从业人员将在其职业生涯中遇到的最大变化。
2. 未来的成功将属于现在已经开始在战略规划上将人工智能和大数据技术纳入投资流程的投资机构。
3. 投资从业人员需要依托协作式的组织文化、认知多样性和 T 型团队，了解并利用这些新技术和应用带来的机会才能取得成功。

本报告中提供的案例可供投资机构和从业人员借鉴，以适应不断变化的投资格局。在向人工智能新时代进发的过程中，这些先锋企业的经验可以成为投资业内的标杆。

¹⁶ 参见例如，www.forbes.com/sites/quora/2017/11/02/will-robots-eventually-replace-humans-as-the-dominant-species-on-earth/ 和 www.barrons.com/articles/will-ai-help-human-investors-or-replace-them-1543071600。

¹⁷ 参见曹实，《2018 亚太金融科技概览》（弗吉尼亚州夏洛茨维尔：CFA Institute，2018 年）和 CFA Institute，《未来投资专业人士》（弗吉尼亚州夏洛茨维尔：CFA Institute，2019 年）。

案例研究

1. 通过机器学习增强交易策略和执行：英仕曼 AHL¹⁸

供稿人：Anthony Ledford

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

英仕曼集团 (Man Group) 是一家主动型投资管理机构，提供多头和另类投资产品，截至 2019 年 6 月 30 日的管理资产规模为 1144 亿美元。英仕曼 AHL 是该集团旗下的量化投资机构，总部设在伦敦，截至 2019 年 6 月 30 日的管理资产规模为 299 亿美元。

英仕曼 AHL 操作种类广泛的对冲基金和多头投资策略，通常涵盖各个资产类别和地理区域。所有业务中还有一定程度的机器学习成分。例如，所有英仕曼 AHL 策略的通用执行流程均使用适应性的智能路径算法，从而选择给定交易的最优市场路径。机器学习在我们的多策略项目中使用最为广泛，使用的方式也多种多样，例如模式识别、趋势跟踪、自然语言处理等。

英仕曼 AHL 早在 2009 年就开始研究机器学习及其在投资领域的应用，但这一尝试并未带来在项目投资组合中的实施。2012 年，我们再次展开研究，不过这次我们使用了不同的机器学习算法。当时团队非常怀疑能否成功，但经过较长的研究测试、模拟交易以及使用英仕曼 AHL 自有资金进行实时交易后¹⁹，第一个机器学习策略最终在 2014 年进入英仕曼 AHL 的项目投资组合。

投资流程

目前，机器学习在英仕曼 AHL 投资管理活动的两个方面影响最大：一是交易策略的开发（这是指生成交易的算法，即买卖什么以及何时买卖），二是提高交易执行的效率（即向金融市场发送并完成交易）。在开发交易策略过程中应用机器学习技术时，我们高度重视如何确保最终策略能够让现有投资组合多元化。换言之，机器学习算法不是执行无方向的搜索，而是经过修改以寻找兼具阿尔法和多元化的特征，减少或忽略我们不希望具有的投资组合特征或无益的效果。这一研究经验的积累过程十分艰难，尤其是在初期，但我们相信现在团队已对此轻车熟路。

在交易策略确定当前最优仓位后（例如做多 700 股 ABC，做空 250 股 XYZ），取得这些仓位所需的任何交易都将发送到执行系统。执行系统的任务是在规定的期限内执行交易以达到要求的各个仓位，同时尽量减少因此发生的交易成本（市场影响）。英仕曼 AHL 运用自己的电子交易执行算法已经有超过十年的历史，并以券商或银行对手方开发的第三方算法为补充。需要回答的关键问题（可能一天需要回答数千次），是完成每笔交易最高效（最低廉）的方式是什么？换言之，我们应该使用什么算法？还是应该由人来处理？

使用机器学习技术可以在线实时回答这一问题，自动将交易分配给某个可用的执行渠道（例如算法 A、算法 B、…、人）。根据我们的经验，人类决策者发现这一最优路径的能力非常不理想，因为交易成本数据不是固定不变的，并且具有极高的噪声。我们从 2016 年底开始研究使用强化学习工具来解决此问题，发现在节约成本和时间方面有很好的效果。从最初少量买卖英仕曼 AHL 的一些期货合约开始，此方面的应用不断升级，现在已用于帮助优化整个英仕曼集团的期货和现货股票委托单流。

¹⁸ 英仕曼 AHL 首席科学家。此案例根据我们对他的采访以及英仕曼 AHL 提供的资料编写。

¹⁹ 在使用英仕曼 AHL 自有资金进行实时交易中监测新策略，是此策略可以进入项目投资组合前需要完成的最终验证步骤之一。

人工智能 / 大数据技术

我们对于全面利用机器学习技术持开放态度；例如，我们不知道是否能够通过深度学习或贝叶斯机器学习算法解决问题。在我看来，我们的任务是探索各种选项，找出哪些有效，哪些不太有效，然后将可增加价值的东西纳入客户的交易过程中。对于已经完成研发并进入客户投资组合的策略，我们已经在使用贝叶斯机器学习、深度学习和模式识别算法进行实时交易。最近，基于 NLP 的策略（基础数据为文章）也已用于实时的客户交易。对这些策略的研究需要使用专业的硬件，这是一种称为图形处理单元（GPU）的处理器，可以在标准计算机 CPU 所需 1/30 的时间内完成深度学习研究需要的计算。

对于交易执行策略的研究，使用简单的时序数据不够充分，即使采样的频率极高，因为相关数据对象更加复杂，例如动态变化的限价委托单（LOB）。与简单时序数据相比，这种市场微结构的研究可能需要指数级提升数据吞吐量和算力，尤其是与机器学习模型结合时，后者本身可能极为消耗算力，即使是较小规模的数据集。

编程语言的选择尤其重要。英仕曼 AHL 从 2011 年开始就在使用 Python 进行交易和执行算法的研究，这些算法在客户投资组合中的实时实施也使用 Python。以前，研究人员混合使用 S-Plus、Matlab 和 R 语言，实时交易实施则使用 C++ 和 Java 语言。研究和实施团队使用同一种语言，可以提高吞吐量，例如不再需要二次编写代码，同时也带来了一些意外的惊喜（例如有利于研究人员和技术人员在同一个代码库中更紧密地合作）。

团队结构和开发过程

英仕曼 AHL 的大多数研究人员都拥有理科背景（例如数学、统计学、计算机科学、物理学、工程学或地质学等等）。而我们机器学习团队的成员则拥有各种非金融背景（例如计算神经科学、在线新闻），并带来在不同领域应用机器学习的博士后研究经验。最初，我们在期货市场应用机器学习，因为这是我们最熟悉的领域。后来，随着我们汲取经验教训，了解哪些有效哪些不太有效，开始逐渐将研究范围扩大到现货股票。英仕曼 AHL 的首席投资官和其他高级管理人员负责确定主要方向；机器学习团队和其他研究人员努力把握这些领域的交易信号，与我们现有的策略形成补充。我们鼓励全体员工就他们怀疑的现行系统的漏洞或可以改进的某些功能提出研究项目建议，落实这一自上而下的战略。

整个团队是高度统一的。研究人员、技术人员和基金经理没有明显的界限；他们都在一个连续体上。成功的研究人员使用的策略最终会应用到客户的投资组合中，他们负责监督其策略的实时交易和执行。

主要经验

过程很艰难。没有现成的经验可以照搬。一个常见的误解是以为将大量的数据、先进的计算机和聪明的人集合起来，就能够提取有用的信号。经验很重要。只有极少比例的数据是“有用的”。另类数据提供了新的机会，但往往不具有足够长的历史。

使用非机器学习工具可以提取到什么，应是决定是否要使用机器学习的出发点。对于使用机器学习取得的额外收获，承担的额外复杂度是否值得？因此，我们的信条是“使用能完成工作的最简单的工具。”

拥抱开源。通过贡献反馈，保持长期参与。形成良性循环。

坚定参与过程。果断决定哪些东西值得追求，放弃看来没有前景的项目。

2. 利用机器学习生成量化模型的信号：纽约人寿投资公司

供稿人：Poul Kristensen²⁰，CFA

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

纽约人寿投资公司是纽约人寿保险公司的资产管理部，截至 2018 年 12 月 31 日管理的资产规模为 3140 亿美元。

其多资产解决方案团队负责管理 100 亿美元的全球宏观资产配置产品。团队成员背景多样，从经济学和金融学到工程学和物理学等等。

投资流程

我们综合使用基本面和量化研究输入。我们的量化研究工具包是我们决策流程的出发点。但对于全球宏观投资，一些市场驱动因素难以量化，因此对前景的基本面 / 定性评估也同样重要。

在量化工具包中，我们使用因子驱动的方法来进行资产配置，也就是将资产类别分解为相应的基础风险因子。我们对这些风险因子进行预测性分析，这基于四种不同的系统性方法：周期、价值、动量和情绪。在这些量化方法中，我们将机器学习技术作为一个重要的输入。方法的设计和因子的选择都基于广泛的模拟、交叉验证以及实践经验。

模型是我们月度投资政策会议上讨论的出发点。月内发生较大的市场转变和数据流时，也会使用人工智能工具。基金经理以及为他们提供支持的量化研究人员都会使用这些工具。

经济周期可能是市场回报的重要决定因素，尤其是在经济加速或减速发生重大转变，或者出现直线下跌和回升时。周期的转变不仅会影响风险资产的平均回报，也会影响更广泛的概率分布特征，例如波动率、偏度（非对称性）和峰度（“肥”尾）。机器学习技术在根据每天传入的广泛数据映射经济周期和“临近预报”市场形态方面作用极大。与传统统计技术相比，机器学习技术还可让我们纳入更多的数据（更广泛的指标）。从而让我们可以获得更可靠的实时周期趋势预测。

人工智能 / 大数据技术

对当前周期形态的预测性分析（例如上行和下行阶段）是作为周期投资策略的输入使用的。对投机级违约率的预测性分析，则在评估信用敏感型资产类别的价值信号时作为输入使用。

一些指标与使用机器学习技术之前相同，但现在我们能够使用和处理更广泛的指标（因此可以压缩指标中包含的信号）。借助机器学习技术，我们可以更可靠、更快速地处理大量指标中包含的信号，从而提升工作流程。

人工智能和机器学习技术不仅适用于投资组合决策本身，同时也越来越广泛影响广义的数据处理方式。人工智能技术可以用于映射特定时间的预测指标并按重要性排序，设置控制面板以高效地监测最关键的指标。使用人工智能和机器学习技术后，创建这种控制面板涉及的工作流程的效率大大提高。

开发过程

包括机器学习工具在内，量化工具包由三至四个团队成员开发和维护。这些团队成员拥有计量经济学、物理学和数据科学、运营研究和工程学等领域的背景。

20 纽约人寿投资公司董事总经理、经济学家、基金经理。此案例根据我们对他的采访以及纽约人寿投资公司提供的资料编写。

他们最初受到各大学研究论文的启发，这些文件记录了机器学习技术在投资策略方面的优势。根据来自这些论文的观点，加上公司内部从业人员的见解，该团队参与了多个项目，探索与整个组织的投资团队协作，在不同的环境下使用人工智能进行投资决策。开发创新的技术十分艰难；现实中也没有教科书式的操作可以借鉴。我们探索了不同的用例，对不同的观点进行试验。与所有创新一样，“试错”是整个过程的自然组成部分。一些技术很早就表现出较好的前景，而另一些技术则没有什么效果，或者过于复杂，难以解释并与真实决策联系起来。然后，研究人员会从头再来。这是一个迭代的过程，预计未来也将继续如此。世上没有免费的午餐。

寻找数据，试验某个想法，以及阅读模型的输出信息时，经常会出现关键的“惊喜”时刻，突然会迸发出某个想法。因此，数据可视化是创造性过程的关键。此外，浏览使用新技术的研究人员的论文草稿和笔记时也不时可能会有“惊喜”。即使用例可能不完全相当，有时仍可能存在有益的见解。

影响 / 主要经验

我们大量使用信用质量、杠杆率、流动性、借款和发行趋势、违约率等指标，作为信用利差和违约预测工具的输入。借助机器学习技术，我们能够纳入更大规模的数据，提高预测的准确性，识别出需要在控制面板中监测的最重要的预测指标。

机器学习技术生成的信号，有利于我们在作出投资组合决策时专注于最重要的指标，从而更好地评估市场中的风险和机会。

机器学习技术生成的信号（尤其是周期信号和价值信号），有利于我们在作出投资组合决策时专注于最重要的指标，从而更好地评估市场中的风险和机会。借助周期框架，我们还能够在衡量投资组合的前景时监测更广泛的指标，节约准备投

资委员会会议以及日常投资组合监测所用的时间。

3. 通过深度学习完善股票成交量预测：道富集团

供稿人：王大军，CFA²¹

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

成交量预测是金融领域中的一个重要问题，尤其是对于提供资产管理或交易执行服务的机构而言。它的重要性在于它不仅有利于机构更好地配置交易资源，同时也让交易员能够更好地了解市场情况。在这一交易项目中，我们尝试了借助深度学习技术来预测成交量，并将此流程提炼成为一种能够按市场预测每日股票成交量的服务。

输入

一家金融机构可能会服务多种投资风格的基金，因此其每日交易行为的时间模式较为复杂，必须通过多种因素进行区分，包括：

1. 金融市场指数。这一类型的因素对于主动型基金尤其重要。主动管理型基金的投资策略通常会依赖将一组金融市场指数作为输入，例如 CBOE 波动率指数或美林 MOVE 指数。因此，这些指数的变动会影响基金的每日成交量。
2. 市场指数成分股调整。这一类型的因素可能会影响被动型基金。被动管理型的基金通常会跟踪某个市场指数，例如标普 500 指数或恒生指数。这些指数的成分股调整一般会导致在生效日的成交量上升。这一现象十分普遍。
3. 历史成交量。历史成交量提供了正常成交量水平的基线和长期趋势。
4. 区域市场时间相关性。一些金融机构在全球经营，包括多个区域市场。即使每个市场都有自己的本地节假日和安排，在一些情况下，这些市场之间仍有相关性。

5. 特殊日期。一些日期（例如星期四、每月最后一个交易日、公众节假日）对日成交量有重大的影响。

为使预测服务的准确度达到综合性金融机构的要求，我们的模型考虑了所有这些因素。该服务每天提取这些因素，返回各个市场的预测成交量数据。如果输入因素的数据累积到一定规模，模型会使用更新的数据进行重新训练，从而满足潜在结构性变化的要求。

人工智能技术

我们预测服务的核心是一种序列到序列模型，它基于一种多层卷积神经网络 (CNN)。该模型能够摄入任意长度的多个时间序列，然后输出预测的序列，以及最终预测。

该模型包含两个编码器网络和一个解码器网络。第一个编码器对市场指数数据进行恰当的转换，然后从中提取特征。第二个编码器对日历指标和历史成交量指标执行同样的操作。这两个编码器的输出将会组合起来，然后输入到解码器网络中。解码器网络使用“注意”机制构建。它包含 CNN (卷积神经网络) 层和扩展 CNN 层。解码器网络中的 CNN 层主要设计用于获取来自特征的短期信息，扩展 CNN 层设计用于获取长期和周期性的信息。

该模型是使用 Facebook 的 PyTorch 深度学习框架构建的。其运行能够使用 CPU 和 GPU 进行模型训练和计算。

团队结构和开发过程

我们的模型开发团队拥有两名数据科学家、一名金融工程师和一名业务分析师。团队成员的教育背景涵盖数学、统计学、机器学习和量化金融等领域。

此项目于 2018 年 8 月启动。模型开发团队一直积极合作，了解业务背景和需求，讨论潜在的特征和数据来源，解决数据问题，确认中间结果。每月与管理层进行评审会议。模型开发人员在这些会议上向有关人员通报最新进展并提供工作示例，有关人员向建模团队提供有关模型增强的建议。

²¹ 道富集团董事总经理。此案例根据我们对他的采访以及道富集团提供的资料编写。

2018年11月，模型的结构固定了下来。建模团队开始将模型打包为一个服务。在此阶段，模型验证团队和IT团队也参与进来。模型验证团队有一名内部模型验证师，负责验证模型在不同场景下的性能。他还就模型的稳健性向模型开发团队提供建议。IT团队有一名架构师、两名开发人员和一名质量保证工程师。他们以安全、有弹性的方式将整个模型重构为一种RESTful服务。然后，用户可以通过多种编程语言（例如C/C++、Java、Python、Matlab）或通过互联网浏览器等任意HTTP客户端访问我们的服务。他们的贡献还包括用户身份验证和授权、模型的弹性部署、加密数据交易、灾难恢复以及审计日志记录等。

如果输入因素的数据累积到一定规模，模型会使用更新的数据进行重新训练，从而满足潜在结构性变化的要求。

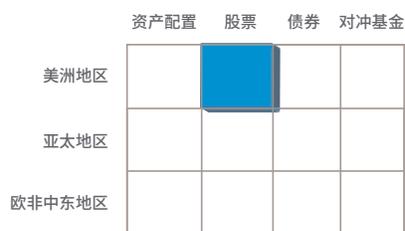
2019年1月，该服务的第一个测试版发布，使用道富的内部交易数据进行了测试。测试数据集包含的每日股票成交量数据涵盖63个市场和各种交易风格的基金。在测试期间，每天将实时数据输入我们的应用程序中，验证其功能和预测准确性。对于美国市场，我们的模型在测试期间的R2达到0.86，其他市场也达到0.54，与预期完全相符，也与训练数据集的矩阵相当。在测试期间，我们的服务还成功预测了富时指数触发的多次成交量突增。

主要经验

我们的成交量预测项目证明了人工智能技术可以成为道富前台核心投资职能的强大补充。它确立了将人工智能辅助模型转化为实用、强大的工具，推动净利润增长的未来项目的标准。未来的项目将在此经验的基础上，进一步提升公司的经营水平，同时遵守有关架构设计、数据治理和项目生命周期管理的标准。

4. 在卖方研究中利用人工智能 / 另类数据分析：高盛

供稿人：Ingrid Tierens, CFA²²
Dan Duggan²³



背景

高盛环球投资研究部数据策略团队与全球的股票和宏观研究分析师直接合作，开展需要利用到分析和量化技能的项目，由研究分析师提供主题性的专业支持并讲述投资用例。过去两年多以来，该团队（常驻纽约和印度班加罗尔）合作发布了近 200 份研究分析报告，涵盖全世界的多个行业和市场。数据集包括应用程序分析、社交媒体、卫星图像、政府数据以及报纸文章等。将这些数据来源与各种先进技术（例如集群和情绪分析）结合，可以得出独特的投资研究见解。例如 2019 年 1 月 17 日，首席股票分析师、CFA 持证人 Jerry Revich 发表了“GS Aggregates Share Tracker: Geospatial Assessment Points to Pricing Upside”，该报告利用公开地理空间数据来作出本地市场份额估计，并将其与本地人口统计信息进行关联。

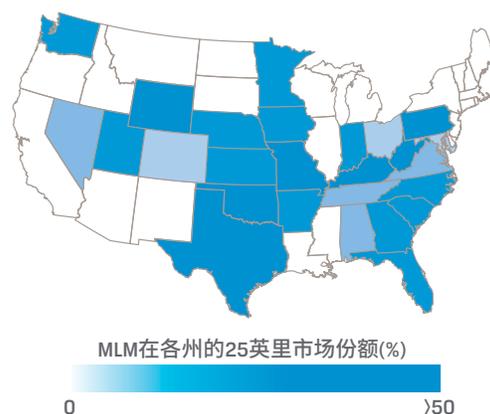
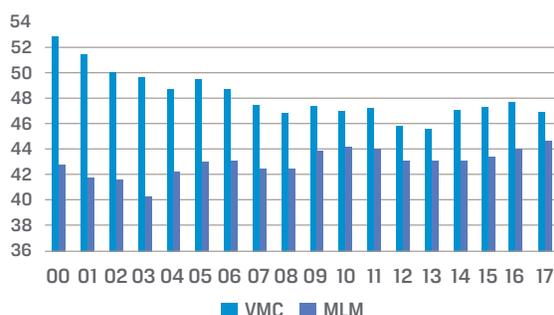
分析方法

“GS Aggregates Share Tracker”回答了与材料行业中这个细分领域有关的下列问题：(1) 如何理解一个高度本地化行业的定位；(2) 如何反映上市公司在一个主要为私营竞争格局的市场的份额；(3) 如何让投资从业人员把握与集料产量有关的季度公司业绩数据的方向性意义。在集料行业（建筑行业混凝土使用的砂、砾石采掘和岩石破碎），产量增长和产品价格利润率是最重要的两个关键业绩指标。市场本身是超本地化的，因此按 25 英里范围来计算的市场份额要比按州或国家水平计算的市场份额更准确。识别一家公司的

矿场位置并分别估计各个矿场的产量，可以深化分析师对一家公司市场地位的了解（例如，公司在哪些市场有定价能力或有重大风险敞口）。通过增加新的数据来源，使用地理空间信息来切实跟踪公司的表现，可以更好地洞察这些公司的经营情况。图 10 复制了 Revich 公司 2019 年财报的附表 21 和附表 8，可视化显示了对 Martin Marietta (MLM) 公司更深入的洞察，按州列举了公司的 25 英里市场份额，并与 Vulcan Materials (VMC) 公司的市场份额进行长期对比。

图 10：MARTIN MARIETTA 的各州市场份额以及与 VULCAN MATERIALS 市场份额的比较

25英里半径内的市场份额（年度时间加权平均值）



资料来源：美国矿山安全与卫生管理局、公司数据、高盛环球投资研究部。

22 高盛环球投资研究部董事总经理。此案例根据我们对她的采访以及高盛提供的资料编写。

23 高盛环球投资研究部副总裁。

预估这一空间的集料产量可以洞察公司的战略：公司专注于哪些市场，潜在并购活动可能会如何改变公司的市场地位，哪些定价能力最为重要。在最基本的层面上，分析用于估计每个季度的内生性增长，并提供有关盈利方向的输入。此外，分析还用于了解市场份额的转变以及对定价的影响，并用来估计公司层面的集料产量加权风险敞口，从而反映公司应对区域性洪灾和飓风等天气事件的能力。

人工智能 / 另类数据分析

“GS Aggregates Share Tracker”将公司数据和公开挖掘数据组合起来，作为模型的输入。它利用了开源地理空间库的优势，还利用了矿场特定的元数据，提供集料产量估计和区位信息。借助迭代式的验证序列，可以对元数据进行更正，减少与派生位置有关的不确定性。通过将 NLP 和公司披露的子公司数据结合，加之分析师的专业技能，从而对实体进行映射。市场份额的计算是位置特定的，因此对全美大约 9000 个矿场分别计算，这些矿场间隔距离 5 至 50 英里，以 5 英里为增量。

团队结构和开发过程

为实现利用另类数据和机器学习，为买方客户提供更好、更有差异化的投资见解的目标，我们努力寻找下面三个职能的交叉点：(1) 发挥主题性的专业技能；(2) 获取所有相关信息，包括“传统”来源和适当的另类数据；以及 (3) 应用先进的分析技术以提取相关见解。

不论项目是针对特定情形，还是长期性的，首先了解我们要解决的投资研究用例至关重要。否则，我们可能会最终浪费大量的时间和精力。研究领域专家和量化分析师之间的沟通十分关键，能够了解研究流程的机会越多，效果将会越好。与研究分析师进行交流，有利于帮助我们确定我们作为一个团队增加的价值是来自嵌入另类数据，提供数据分析专业支持，还是帮助分析师将他们的见解系统化到融合到数据产品中。有时，我们发现我们要参与的项目并不适合。

量化分析研究团队的建模流程与分析师团队完全统一，这两个团队从始至终全程参与模型的构建和验证。对于“高盛集料份额追踪”，我们的量化分析团队与 Jerry Revich 领导的应用分析师团队合作，全程管理整个流程的各个阶段：根据各公司的重大事件，为每家公司迭代实体映射结果并交叉检查公司矿场位置模型映射、历史平均产量结果以及特定的验证顺序。最初的模型仅依赖上市公司数据，但随着模型复杂性的上升，它扩展包含了全部 9000 多个美国矿场，为分析师团队创造了极大的范围扩展空间。

主要经验

基于我们的经验的一些建议：

- 不要低估更先进的技术和方法在提高投资从业人员效率，从而将他们解放出来，使他們有更多的时间创建更好的阿尔法见解的潜力！
- 另类数据不一定会成为系统性策略投资经理的优势。事实上，我们认为更小众、行业特定的数据集本身更适合基本面分析师或深入探究特定资产的基金经理，而不适合涵盖广泛资产的分析师或基金经理。
- 根据我们的经验，单个数据集或单个方法从来都不是研究产品的唯一驱动因素。对于我们而言，另类数据是补充，但本身并不是目标。
- 好的研究分析师往往会注意与他们所研究资产相关的数据，但可能不知道如何最好地访问和 / 或分析这些数据。
- 另类数据的利用不一定意味着要算无遗策。有许多公开数据可用，如果与已经关注的其他数据来源恰当结合，可能会大有裨益。

5. 借助人工智能和大数据剖析财报电话会议： 美国世纪投资公司（AMERICAN CENTURY）

供稿人：Tal Sansani²⁴

	资产配置	股票	债券	对冲基金
美洲地区		■		
亚太地区				
欧非中东地区				

背景

美国世纪投资公司于 1958 年在堪萨斯城成立，为金融中介、机构投资者和个人投资者服务，目前的管理资产规模超过 1600 亿美元。公司下属的股票量化投资团队由研究、投资组合管理以及投资技术方面的专业人士组成。

股票量化投资团队在硅谷的中心区域办公，在将人工智能技术纳入投资流程方面，已经取得了很大进展。该团队管理大约 140 亿美元的资产，他们使用的流程融合了人工智能和大数据技术的元素。

投资流程

美国世纪投资公司的选股模型是再平衡流程中一个直接输入的部分，此外还包括大量风险因素、交易成本因素以及许多其他考虑因素。基金经理在审核交易列表时，会敏锐地关注名称层面的风险、新出现的宏观消息以及模型输出的变化等。

选股模型的目的是以系统性的方式识别未来回报的基本面驱动因素。它基于四大支柱：质量、增长、估值和情绪。每个支柱都由大约五个专有因素组成。模型预测会随着新信息的输入持续更新。

情绪支柱旨在使用未在公司最近财务报表中充分反映的信息，来捕捉公司前景会出现的根本改变。例如，我们的模型会收集来自做空机构的信息，识别出顺应产品市场和地域趋势的公司，同时系统性地避开那些表现出过热迹象 / 羊群效应的个股。

我们情绪模型的一个关键成分用于分析管理层在季度电话会议上的语言。我们的电话会议欺骗模型由四个部分组成：遗漏（未披露关键细节），浮夸（管理层夸大其词、过度粉饰），混淆（过度复杂地讲故事）和指责（逃避责任）。为避免偏见，我们会考虑给定管理层团队的独有风格和整个行业可比公司的集体语言。

人工智能 / 大数据技术

我们的机器阅读模型高度依赖 NLP 技术。在此领域，我们利用了各种情绪分析技术、文档分类技术、词性标记技术、实体识别技术和主题检测技术。许多此类语言处理技术都需要通过机器学习来准确高效地执行。

此外还需要多种技术来抓取、清理混乱 / 非结构化的数据并将之结构化，尽管这些操作不那么引人注目。例如，网站上的数据需要使用 HTML/XML 语义分析工具、字符串模式识别工具，以找出特定内容中的关键词，此外还需要系统性地剔除噪声（例如广告或不相关的链接）。比以往需要更多的是，流程中还涉及到多种语言的分析。需要的工具则多种多样，从已经存在数十年的技术（例如正则表达式和字符串距离函数），到快速发展的 Python 和 R 开源平台，再到强大且有针对性的第三方科技公司的服务。

数据集越大，需要预处理的数据越多，需要的计算资源也越多。虽然我们的团队通常会利用本地管理的高性能计算集群，但更大型的任务则需要使用 Amazon Web Services 或 Google Cloud 等弹性的云服务资源。

24 美国世纪投资公司副总裁兼高级量化研究员。此案例根据我们对他的采访以及美国世纪投资公司提供的资料编写。

团队结构和开发过程

该团队的最佳想法（使用或不使用人工智能）常常始于与基金经理和研究人员的交流以及他们的意见，以合作创造独特且经济上合理的选股观点。我们的目标，是将基本面的洞见转化为系统性、客观且可重复的算法。

一旦观点得到定性核查，研究人员会与投资技术人员一起（通常是一位研究员和一位技术人员）采集并处理必要的输入数据，并在必要的技术和方法上进行协作。

在此案例中，我负责领导研究和开发工作，在团队中兼任研究员和投资技术人员的职能。在此过程中，我们会征求团队中许多人员的意见——在整个研发过程中，首席投资官、研究人员和基金经理都提供了反馈和见解。技术人员承担后端设计决策的协调执行职能。

我的学术背景为应用数学和计算机科学，我获得的大部分金融和技术训练都属于在职学习，包括向同事和导师学习，以及自学。

不论是从研究还是从技术的角度，开发过程都具有极高的迭代性质。我首先利用金融期刊和心理学书籍建立了假设基线，概括了与欺骗有关的语言学模式类型。随着研究的深入，我征求了团队中研究人员和基金经理的反馈意见，例如，询问他们“这与你在参加电话会议时期望听到的是否一致？如何改进或完善此模型？”其中一种增强一项改进涉及根据相关行业和 / 或管理层的具体性格（例如是一位年轻且过度乐观的技术型首席执行官，还是有着石油开采业丰富从业经验的首席执行官）对模型进行语境化处理。

最成功的投资团队，是通过与技术人员合作来提供经济上稳健且进行周密实施的解决方案。

此模型的开发要求采取多学科的研究方法。这包括学习心理学书籍，以研究来自各种上下文和人（例如儿童、罪犯、企业高管等）的欺骗信号。当我们从日常生活中的骗子以及有所掩饰的公司高管身上分析出清晰、共同的语言模式，这就是“恍然大悟”的情形。

主要经验

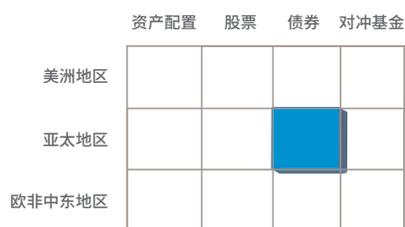
- 首先关注投资问题。
我们的数据科学家拥有综合的技术型背景，计算机科学和统计学背景最为常见。值得指出的是，虽然这些背景提供了较好的基础，但如果没有股票投资领域的专业知识，也很难加以应用。最成功的投资团队，是通过与技术人员合作来提供经济上稳健且进行周密实施的解决方案。
- 机器的智能程度取决于它所学习的数据。

取得额外的数据往往比构建更复杂详尽的模型更为有效。用于训练的数据越综合全面，机器对新事件的处理结果将会越具有概括性，从而减少过拟合等常见的陷阱。利用更加丰富多样的数据，对于股票投资尤其重要，因为与估值相关的信息超出了传统财务报表的范畴。

- 人和机器是互补的力量。
人类分析师擅长就相对少量的公司做详细分析与研究，而机器人分析师可以一次对上千家公司系统性地应用其客观、独特的见解。我们认为，能将这两者力量发挥好的团队将会在竞争激烈的环境中处于优势地位。

6. 人工智能和大数据助力信用资产投资组合管理：中国人寿资产管理公司与中证信用增进股份有限公司

供稿人：杨融²⁵
曲靖²⁶
范思雯²⁷
田巧梅²⁸



背景

中国人寿资产管理股份有限公司（简称“国寿资产”）是中国最大的资产管理机构，涉足国内外公开市场、公募（共同）基金、另类投资等多个领域。截至 2018 年底，国寿资产的管理资产规模超过 3 万亿元人民币（4360 亿美元），主要为固定收益投资组合投资。因此，构建技术解决方案对于国寿资产尤其重要。

中证信用增进股份有限公司（简称“中证信用”）是一家信用科技综合服务商，总部位于中国深圳。CreditMaster 是公司面向中国债券市场推出的综合信用增进解决方案。国寿资产从 2018 年 7 月开始使用 CreditMaster。

工作方法

中国境内债券市场在短短十几年里取得快速发展，目前规模已达 13 万亿美元²⁹，但高速增长的背后也面临着诸如信息披露不完善等问题。在国寿资产的投资组合中，信托投资等“非标产品”的规模已经增长至 3300 亿元人民币的规模。国寿资产希望能够借助一种技术解决方案，在投资流程中对固定收益产品进行估值，同时监测整个市场和单次发行的动态。

CreditMaster 具有满足国寿资产需求的多种功能：

- 信用风险数据（即数据超市）包含围绕信用主体、金融工具、风险预警、信用评估引擎、评估历史以及用户账户 / 访问权限组织的数据。
- 评估引擎包含了来自评级机构或其他第三方的评级数据，可以根据上面提到的数据和其他数据实时生成自定义的评级。
- 分析模块将 CreditMaster 纳入信用分析师和风险经理的工作流程。该模块还允许客户根据自己的分析和经验调整模块参数。
- 通过自动化报告功能，客户能够使用 CreditMaster 中的可用信息自动生成标准化的信用报告，然后由分析师加入自己的结论，完成报告。

此解决方案可供国寿资产的信用分析师、固定收益组合经理以及合规团队使用，用于信用分析和投资监督流程。

在信用分析和投资流程中，系统将会根据内外部数据以及模型和评级自动完成大约 80% 的分析。公开信息包括公司年报 / 财务报表、法律诉讼、监管备案和新闻等，例如有关 20000 多家股票和债券发行人及相关金融机构的信息，以及 1700 个地方政府和地方企业债券发行人的财务和经营信息。

国寿资产的分析师和基金经理在预测任何特定发行的违约概率时运用的是逻辑回归模型。系统还将根据当前和目标投资组合的风险收益特征做出发行推荐。系统还会自动生成预警。基金经理可以根据自己的分析，决定对预警名单上的特定发行采取什么操作。

25 中国人寿资产管理公司固定收益部董事总经理。此案例根据我们对供稿人的采访以及中国人寿资产管理公司与中证信用增进股份有限公司提交的资料编写。

26 中证信用大企业信用增进业务部总经理。

27 中证信用战略与发展副总裁。

28 中证信用高级产品经理。

29 参见 www.cnbc.com/2019/04/01/china-bonds-debut-on-bloomberg-barclays-global-aggregate-index.html。

CreditMaster 通过整合内外部信息和建模渠道，节省了信用分析师在系统之间切换的时间和精力。在对发行人评级进行审核时，国寿资产的分析师不再需要打印不同来源的数据和文件，而是直接从系统中可以看到。这样可让分析师专注于进行分析并得出结论。

同样，风险经理不再需要翻阅多个网站以跟踪有关发行人的新闻和社交媒体发布内容。从公共网站和社交媒体采集的数据和新闻，以及从第三方供应商购买的数据，都整合到了系统中，并在触及信号时生成有关发行人的预警。

人工智能 / 大数据技术

系统利用人工智能和机器学习模型，向投资从业人员提供收集和监控综合风险信息的工具，包括：

- 分布式的多源数据接入系统，以秒级检查大量公开可获取的上游数据，涉及数百万个数据采集点；
- 自然语言处理（NLP），从动态的非结构化文本数据中获取有用信息，以进行分析和建模（例如使用 SimHash 算法生成的指纹信息可以帮助筛选同样的主题信息，保留大约 30% 的相关信息）；
- 双向长短时记忆（BiLSTM）模型，对发行人和产品执行指定的实体识别分析，从而确定是否有文章关注它们或只是提到了它们；
- 文本卷积神经网络，对文本样本进行情绪分析，在识别负面信息方面的准确率高达 90% 以上（对于预警系统极为有用，可以向团队发送有关投资组合中持仓的相关消息）；
- 支持中证信用建立覆盖 7000 万家公司网络的知识图谱（数亿个股票所有权和企业任命数据点），让系统能够进一步利用知识推理技术并对关联方信息执行深入挖掘（这对于被监控实体的关联公司发生风险事件时防止风险蔓延十分有用）。

团队结构和开发过程

中证信用的团队承担从信用建模到系统开发的多个职能。主要职能如下：

- 信用分析师负责开发和维护中证信用行业信用模型的方法论，为客户自有的信用模型提供备选方案。
- 量化专家使用量化和机器学习技术来开发和维持中证信用的行业信用模型。
- 数据科学家使用各种人工智能技术（例如 NLP 和文本卷积神经网络）来采集和处理来自公开来源和第三方供应商的数据，从而生成信号。
- 工程师是目前人数最多的组成部分，负责从前端到后端的系统开发工作，包括用户界面和数据库开发。
- 产品经理负责向工程师以及量化专家和数据科学家沟通客户需求。

中证信用特别看重不仅擅于履行自己职能，而且可以延伸到一两个其他职能领域的专业人才。

主要经验

一个综合性系统的开发需要不同的技能组合，包括对不同数据集的标准化和清理，到利用这些数据集生成投资见解。中证信用发现，将上述五大职能集合于一个团队中，向着一个目标努力，是很重要的。传统的模式将数据科学家、量化专家和系统开发分为不同的部门，不能发挥这样的效果。

中证信用服务产品的重大版本发布需要国寿资产管理层审核确认，在重大版本发布的间隙，中证信用的团队会每月向终端用户提供产品“迭代”。中证信用的团队发现，这种方法对于及时收到客户的反馈极为有用，减少了整个组织中投资和技术部门间发生沟通误解的风险。

7. 运用人工智能和大数据技术受理和处理保险索赔以及评估企业风险： 平安集团

供稿人：肖京³⁰

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

平安集团成立于 1988 年，现已成为全球最大的保险集团，拥有 180 万名员工。平安集团致力于建设科技型个人金融服务集团，业务涵盖保险、银行和投资等领域。借助“ABCDs”技术（人工智能、区块链、云计算、大数据和信息安全），平安集团还建立了五大生态系统的综合能力：金融服务、医疗健康、汽车服务、房产服务和智慧城市服务。

平安集团拥有丰富的优质金融数据。从 2015 年第一次成立不到 10 人的人工智能团队以来，我们看到大数据在满足用户个性化需求和帮助平安实现业务增长方面具有巨大潜力。四年后，团队增加到 1000 人，覆盖金融、医疗健康和智慧城市等多个业务线，处理多种应用场景，例如风险控制、欺诈检测、智能医疗、医学诊断和治疗、运营优化、智能金融和精准营销等。在本文中我们将重点介绍“人工智能 + 闪电理赔”平台。

业务流程

平安智能闪赔通过图像识别自动完成车损识别，每天处理上万件理赔案件。客户只需要上传几张受损车辆的照片并在事故理赔网站上申请理赔，系统将在几秒钟内自动处理申请并精确计算赔付金额。该系统涵盖一系列关键模块，包括照片质量评估、保险车辆核实、汽车零部件区分、各部位的损伤识别、赔付计算以及欺诈检测等。

该系统已在平安集团运行一年多，每天成功处理超过 3 万件理赔。它不仅提高了理赔处理的效率，改善了客户体验，同时也防止了高达数十亿元的潜在欺诈。目前该系统已经通过平安金融壹账通平台向整个保险行业开放。平安集团每年处理超过 1100 万件理赔，98.7% 在一天内赔付，自助服务率达到 60%。

另一个应用是企业风险评估。中国有 7000 多万家注册企业。企业信息主要来源于三个渠道：工商注册和日常经营、公告发布（包括新闻和社交媒体）以及业务关系（例如供应链、投资、法务、高管等等）。为了组织和分析这种丰富而动态变化的数据，平安集团开发了企业知识图谱应用欧拉图谱（Olatop）。

结合知识图谱、NLP 和机器学习等先进的人工智能技术，欧拉图谱可以精准定位并解决传统企业运营分析框架的缺陷和难点问题。它是完全依托金融分析领域研发出来的，具有成熟的工具、完善的功能和定制的解决方案，包括财务分析、债券分析、情绪分析、产业链分析和政策分析等功能。所有功能都可在四个主要领域实现：风险管理、精准营销、智能投资和公共服务。

人工智能 / 大数据技术

为了支持智能闪赔平台的功能，我们开发了一些人工智能技术，例如图像处理和区分、对象检测和识别、知识图谱以及异常分类等。

在欧拉图谱中，数百万的法律诉讼可以自动进行解读，超过 4000 万个诉讼管辖已经提取并纳入。来自 300 多个新闻和社交网站的企业相关信号也会自动收集（每天数十万篇文章）并每隔 10 分钟更新一次。然后，深度图谱分析算法进行开发，以支持风险评估等业务决策。

30 平安集团首席科学家，也是平安技术研究院人工智能中心负责人。此案例根据我们对他的采访以及平安集团提供的资料编写。

业务逻辑也已直接集成在欧拉图谱中。例如，风险事件针对投资、债券和贷款设计了不同的业务逻辑，并通过对社交媒体和新闻媒体数据的情绪分析提取相关信号。此外，不同的上下游关系量级也作为风险指标进行编码。当上游出现风险事件时，将在整个图谱网络传递影响并按照关系量级进行消减处理。

团队结构和开发过程

平安技术研究院人工智能中心分为业务应用团队和技术应用团队。在所有的团队中，有两个团队是协作来交付基于知识图谱的智能解决方案：一个团队负责大数据分析、数据建模、通用工具开发等等，另一个团队负责知识提取、文本分析和图谱构建等。前者负责清理和整合结构化和非结构化数据，然后后者使用其知识提取和图谱构建能力交付最终解决方案。

目前，我们的团队拥有超过 1000 名人工智能专家，大部分都来自国内外顶尖大学，例如卡耐基梅隆大学、哈佛大学、北京大学、斯坦福大学、清华大学、牛津大学和中国科学技术大学。团队的核心成员都是计算机视觉、NLP、医学影像处理和数据分析等领域的技术领军人物，平均工作经验超过 20 年。加入平安集团前，他们在顶尖高科技公司担任重要职位，例如谷歌、微软、IBM 和优步等等。

对于开发过程，我们受到的启发通常都来自大众业务的许多痛点，促使业务专家和人工智能科学家通过头脑风暴寻找对应的技术解决方案。每个参与者都贡献自己的专业技能和见解，团队最终达成共识，制定实施计划。

该系统已在平安集团运行一年多，每天成功处理超过 3 万件理赔。它不仅提高了理赔处理的效率，改善了客户体验，同时也防止了高达数十亿元的潜在欺诈。

平安集团的人工智能开发经历了四个阶段，包括建设基本认知能力的起步阶段、构建知识图谱系统的学习阶段、结合业务场景输出智能解决方案的专业阶段和挑战人类智慧最困难领域的创造阶段。在每个阶段，平安集团都采用灵活敏锐的开发方法，并通过设置明确的目标、工作计划和优先级，在快速产品迭代和定期回顾的帮助下为企业带来价值。

8. 情绪分析：彭博

供稿人：Gary Kazantsev³¹

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

彭博成立于 1981 年，通过彭博终端及其企业产品和数据推送服务，向金融机构和组织提供金融软件工具，例如证券交易平台、数据服务、新闻资讯和分析等。公司在十多年前开始投资机器学习和人工智能应用，通过自然语言理解和问答以及信息提取和创新的新闻与分析应用程序，覆盖从提高客户服务体验到改善彭博终端用户体验的广泛范围。

最早期有一项应用是情绪分析产品，从 2009 年开始就已向客户开放。之后，产品不断升级，从英语扩展到其他语言版本（例如日语），从股票扩展到大宗商品和外汇市场，从新闻扩展到社交媒体（例如 Twitter）等其他内容类型。现在，该产品还可以处理其他问题，例如命名实体识别和消歧、主题聚类、主题检测、市场影响分析等等。尽管机器学习在此领域的应用可追溯至 2009 年，它现在已被全球资本市场的重要参与者用于多种应用，包括从风险分析和投资组合构建到产生阿尔法的广泛领域。

如今，彭博提供多种机器学习支持的先进分析工具，该产品仅仅是其中一员。

工作方法

学术文献和行业已经探索了多种情绪分析的方法。2009 年开始开发此产品时，我们选择使用最先进的技术。为此，我们使用了有监督的机器学习并开发了创新的集成构建方法。

文本分析的关键问题在于选择目标变量，也就是精准设计我们希望机器回答的问题。这一问题存在许多可能的视角。例如，一些人希望评估文件作者的内心状态或意见。有些应用场景是与这一视角有关的。但在金融领域，这会带来额外的模糊和复杂性，让本已困难的问题雪上加霜。与此相反，我们选择预测读者的意见，也就是新闻故事的消费者，市场的参与者，我们选择要问的问题，大致是如果您是标的公司的多头投资者，当您在没有其他信息的情况下读到该故事，您会将它视为利好、利空还是中性？这种视角的变化以及机器学习算法的目标变量，会在实践中造成很大差异。

这些模型的训练数据是在内部收集的，我们对拥有金融背景的注释员进行了培训。采样方面需要特别注意，以确保数据中代表的来源、文件长度、主题、市值、行业、财报期间等有充分的多样性。对于模型开发十分重要的关键指标是注释员之间的协议，即对于所设置问题的答案，人类注释员达成一致的比率是多少？如果一致率较低（例如偶然一致），则问题的设置不佳。如果一致率高，则我们很可能会构建一个机器，以可靠地复制人类判断力的这一方面。

使用这种谨慎选择出的问题，我们可以在新闻方面达到 80%-92% 的一致率。而对于 Twitter 等社交媒体，每种语言理解任务的一致率一般较低。这样的数据自然就是我们所构建的任何分类的固有准确度边界。需要指出的是，词汇（词典）等简单的方法无法按照要求的准确度来解决这一复杂而模糊的问题。我们最终构建了更为复杂的模型（非线性支持向量机、包含复杂语言功能的模型以及最近的循环神经网络），并使用数据驱动的方法将它们纳入整合流程。最终的机器学习整合属于实时的系统，每天处理超过 200 万份文件，每份文件的处理只需数十毫秒即可完成。然后将每个故事和实体的情绪数据（例如对于一份同时提到谷歌和微软的文件，您可能看到对谷歌的情绪较为正面，概率为 75%，而对微软的情绪则较为负面，概率为 50%）汇总到一个置信度加权的每一证券时间序列或指数中，以便在投资组合构建、策略制定或风险分析中使用。

31 彭博量化分析技术策略主管。此案例根据我们对他的采访以及彭博提供的资料编写。

大量客户早已开始使用情绪分析数据，包括在彭博终端上以及通过我们的企业事件驱动信息推送 (EDF) 产品使用。通过彭博终端上的一些功能，客户可以进行数据可视化操作，将其与其他变量关联，设置预警，或者使用它们来监测投资组合。情绪数据还可以用于自动化生成故事。在企业版中，情绪数据一直是投资组合构建和优化、风险分析、不同周期和不同持仓期限的交易策略制定、执行以及统计套利的一个因素。多年来，情绪数据消费者的构成也发生了变化。最初，这一“另类”数据的主要用户是量化对冲基金和做市商，而今天，大多数机构投资者都在使用它或者计划将它纳入其工作流程。

人工智能 / 大数据技术

命名实体识别和消歧 (NER/NED)、主题分类、情绪分析以及聚类，都是现代机器学习技术在 NLP 领域应用的例子。这些工具可以在整个行业更广泛地使用，以了解非结构化数据，例如即时聊天、电子邮件、分析报告、季度申报和年度报告、合规说明或其他交易相关信息。例如，许多云服务供应商都提供可公开使用的 API，因此现在能够有效地将财报电话会议的音频流处理为转录文本。一旦数据变为文本形式，就可以使用这些工具来构建集成的市场影响模型，随着电话会议的进行持续更新。

团队结构和开发过程

支持并助力这些高级分析技术的整体系统，是一种复杂的大型高科技解决方案，可让客户执行多种其他的工作流程，例如搜索、预警、研究等等。十多年来，成百上千的人参与了它的研发。现在，彭博的人工智能和数据科学工程部门已经拥有 200 多名员工，他们具有软件工程和高性能分布式系统、数学、计算语言学、机器学习和自然语言处理领域的资深背景。

这些产品的开发是一件高度以客户为中心的活动，我们与客户紧密合作，设计和构建解决这些问题的方案。例如，只有两个人的小工程团队为了响应客户的请求，开始构建情绪分析方案。在开发的最初六个月内成功演示初步原型后，团队规模不断壮大和扩展，以向客户交付产品并将技术延伸到问题的其他方面。

十年之前，机器学习还是天方夜谭。计算语言学、自然语言理解和机器学习的工具和文库也都更为初级。因此大多数的开发都不得不在内部进行。开发过程强调需要建立分布式计算环境以及根据金融领域进行调整的自然语言理解工具，这促进了该领域中多年的大力投资。正是得益于基础设施和数据流程（例如分布式计算、GPU 硬件）方面的投资，团队现在才能够以远远更快的速度执行这些及类似的项目，极大地缩短了将新解决方案推向市场的周期。

主要经验

情绪分析产品开发过程和随后的广义文本分析项目提供了多项重要经验。正如前面所总结，端到端的评估十分重要。谨慎选择目标变量也非常关键。从简单的基线出发，在模型开发中的每次迭代都执行错误分析十分重要；否则我们可能很容易发现某个特定的日期或季度本身是有利的或不利的——这可能归因于（例如）底层数据的固有偏差。考虑到这个原因和其他，可解释的模型和人机回圈对我们来说非常重要。

9. 建设数据科学团队：施罗德

供稿人：Ben Wicks³²
Mark Ainsworth³³

	资产配置	股票	债券	对冲基金
美洲地区				
亚太地区				
欧非中东地区				

背景

施罗德数据分析部（Data Insights Unit）由 Ben Wicks 和 Mark Ainsworth 组建于 2014 年。在与投资部门负责人交流中，Ben 被要求考虑施罗德如何通过投资来获得数据驱动研究领域的竞争优势，并领先获得针对单个公司的差异化洞察。

用例：投资观点的产生

在联合国发布有关智慧城市主题的报告后，一位基金经理决定搜索可能受益于此主题以及全球城市化率上升的公司。

第一步是搜索提到“智慧城市”和“未来”等词语的文章，结果有数千条匹配的文章。然后我们运用了多种机器学习技术：

- 使用一组 NLP 算法提取文章中提到的情绪、概念、主题、实体以及关键词。关键是，我们还识别了任何提到的法人实体，包括“主公司”。
- 下一步是使用某种降维方法来进一步简化。我们使用的具体算法使用力导向图，图中代表新文章的点在二维平面上放置，然后迭代相互“拉”近相似的文章，并相互推开不相似的文章。随着时间推移，结果是讲述同一种观点的文章将集合到一起。
- 以这种方式放置后，使用无监督的机器学习算法将文章分配成簇。

最终结果是一张可视的图谱，在屏幕上放置海量的新文章，类似的文章组合在一起，并根据所分配的簇使用颜色代码标注。

簇的特征以文章所提及的公司名称来表示，“主公司”作为文本附加到图谱中。对此“新闻图谱”进行目视检查发现了多个有趣的特征。尤其是在图谱的边缘有一个明显的文章簇，提到了基金经理以前没有听说过的一家公司。

在被提醒基金可能会对这家几乎无人研究的小公司进行投资后，基金经理对该公司的业务进行了财务分析，并确定其基本面和前景都不错。随后，该公司的股票被添加到投资组合中，大约两年后，它为基金带来了 25% 的回报。

团队结构和开发过程

当时，Ben 是全球和国际股票团队的一名基金经理。他认为现代数据科学技术和新的数据来源其实可以作为一种资源，以补充现有的传统股市分析，从而建立面向长期投资，并且与同业公司相比，拥有竞争优势的有力组合。

金融业并不像学术界和其他行业一样时刻与数据科学和分析的发展保持同步，因此需要吸引来自行业之外的高级管理人才。Mark Ainsworth 拥有 20 年的数据科学和分析经验，曾在一级方程式赛车、电信和零售等公司和行业工作。Ben 相信，Mark 掌握的技术可以帮助资产管理机构为基本面型投资者提供竞争优势。

Mark 于 2014 年加入施罗德，他与 Ben 一起成立了该团队。通过多个成功的概念验证证明自己的潜力后，他开始引入数据顾问和数据科学家以提升团队的能力。来自内部 IT 部门的几位数据工程师被组合在一起，帮助搭建数据管道需要的基础设施。同时，Mark、Ben 以及开发团队开始引入新的数据集，找出更多用例。

32 施罗德数据分析部洞察和研究创新负责人。此案例根据我们对他的采访以及施罗德提供的资料编写。

33 施罗德数据分析部洞察和分析负责人。

团队聘用的数据顾问和数据科学家背景多样。一些人来自拥有良好的大型数据科学和分析团队的公司，曾负责处理专有数据，以帮助制定有关公司方向的决策。另一些人则来自数据科学技能是必备条件的学术或科研实验室。此外，团队从成立开始就招募了应届大学毕业生。

开发过程

对数据科学职能这一理念感兴趣的投资者也对引入的数据集种类以及团队工作的项目类型有所影响。在此过程的早期阶段，一个关键的部分是一些投资从业人员与数据科学家之间定期举行的头脑风暴会议，这提供了极为宝贵的机会来产生和测试投资观点，找出多个投资团队共同感兴趣的领域。

在最初的几年里，我们也吸取到了一些宝贵的教训，关于如何将这样的团队整合为一个基本面型的资产管理团队。企业中的每个人都是不同的。一些人从一开始就喜欢团队的理念以及团队未来能够做到的事情，但并非所有人都饱含热情。让大家团结在一起的，是大家对看到结果和收获的期许。在提供有价值的洞察后，我们在内部对于团队可以做的事情，甚至收到了更积极的反馈。随之而来的，是更多希望参与到这一项目中的投资者纷纷就他们希望看到的数据集或他们认为有用的工作提供反馈。

所有投资人员都使用过来自数据洞察部的成果——不论是为某个分析师定制的内容，或者是自动化工具生成的报告，及我们使用大型数据集和机器学习算法做出的定期报告，这些都发送到了他们的收件箱。从个股分析到使用广泛的宏观数据，各类投资者将这些成果用于多种目的。

人工智能与大数据技术

团队工作的一个关键要素始终是所用的技术；团队的作用可以简单概括为，从无法使用 Excel 进行分析的数据集中获取洞察（毕竟，如果可以用 Excel 完成，投资分析师完全能够自己搞定）。因此，团队的工作是以特定的工具和技术为中心。

一个工具是大数据——巨型数据集往往含有数百亿个数据点，无法通过单台计算机处理。对于这些数据集，团队将云技术（包括 Amazon Redshift 等大规模并行计算的数据仓库工具）和我们使用 Hive 数据库的本地 Hadoop 集群结合。

另一个工具是地理空间数据，此类数据仅在地图上才有意义。为此，团队使用了多种专业工具，例如 QGIS（一种流行的开源地理信息系统工具，用于查看和生成地图）和 PostGIS（一种储存和处理此类数据的强大数据库）。

团队还使用机器学习和贝叶斯推理算法等强大的技术，从而掌握数据中的模式并且确保可预测性。对于这些，我们使用了强大的 GPU 支持的机器以及 Hadoop 集群上的 Spark。

目前，团队使用的主要语言为 R 和 Python。我们使用 Kubernetes 和 Docker 将构建的工具部署到生产环境中。Tableau 是一种极为常用的工具，可以简化可视化任务，当然我们也使用 R Shiny 和 Dash 来构建分析类的可视化界面。与所有此类工具一样，一个重要的任务是剔除任何干扰，提供明确的相关信息，以便用户可以看到洞察内容。

机器学习是我们使用的一个强大工具。现在，我们投资团队的同事已在经常使用多种机器学习支持的工具。贝叶斯算法最有效的用途之一，是在跟踪品牌氛围的时间序列中，识别出何时发生了显著变化，然后标记出来，提醒投资团队的同事。研究分析师和基金经理可以将此信息与他们所知道的公司其他相关信息和价格进行结合，判断他们是否应该调整投资组合仓位。

主要经验

我们得到的两个主要经验是：(1) 要有真正相信此类创新的高管提供支持；(2) 团队要拥有正确的技能组合。团队的建设需要有来自行业之外的专家，此外还需要有业内专家，清楚了解哪些种类的见解是有价值的。

建设数据科学能力听起来让人振奋，也非常时髦，但如果没有积极、有益的结果，则完全没有意义。在初期阶段快速取得的一些胜利，帮助我们证明了团队的价值。数据科学能力不能是刻意为之的时髦游戏；它需要为企业增加价值，不论是带来另类数据还是构建机器学习算法来发现数据集模式中的显著变化。

团队始终对于我们在做什么以及我们可以做什么持开放态度，并且在非常早期的阶段就已决定让团队成为一个对所有投资者开放的集中化部门。这意味着没有人担心使用数据或团队的技能花费的成本，或者他们使用了他们并不打算用的资源。最为关键的是，这一决定也确定团队的责任是寻找可以覆盖多个投资团队的价值所在，从而创造单个投资团队无力承担的价值。

来自大型数据集的信号的最佳“用武之地”在于，帮助构建对公司总体状况以及 / 或公司经营环境的了解，而不是在不考虑其他因素的情况下孤立地进行交易。

10. 专题焦点：通过机器学习增强 MPT 的有效边界

供稿人：Marcos López de Prado³⁴

背景

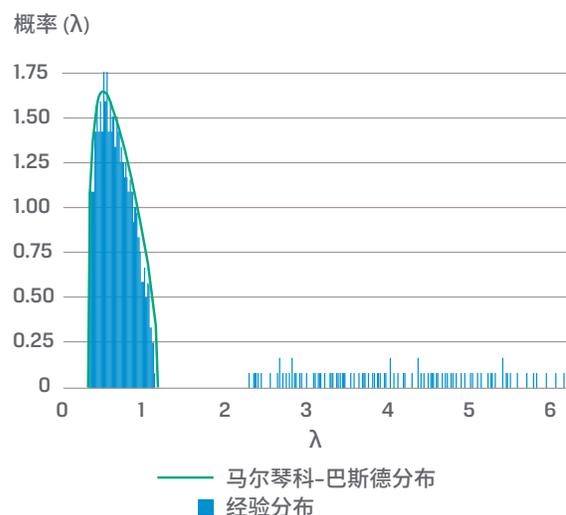
从 20 世纪 50 年代开始，马科维茨的现代投资组合理论（MPT）一直是资产配置领域应用最广泛的方法，带动了许多延伸理论的出现，包括布莱克 - 利特曼模型。³⁵ 但在实践中，这些凸优化解决方案的表现极差，以致于完全抵消了分散化的效益。³⁶ 例如，在金融应用领域，大家都知道样本中最优化的投资组合在样本之外应用时，往往会连基础的（平均权重）配置都跑不赢。³⁷ 对这种落后表现的解释是，有两方面的原因导致了这些最优化的投资组合无法进行稳健估计：噪声导致的失稳和信号导致的失稳。López de Prado 在本文中为我们讲述了机器学习技术如何解决一般意义上的凸优化陷阱，尤其是 MPT 的陷阱。

噪声导致的投资组合失稳

财务相关性指标无法进行稳健预估。即使是基于因子的相关性指标（例如 Barra 模型中的指标），其信噪比也极低。其原因是我们至少需要 $\frac{1}{2} N(N+1)$ 个独立观察点来确保协方差矩阵不会调整不佳。例如，一个只有 100 个工具的较小相关性矩阵，要求至少 5050 个独立且均匀分布的观察点，或者 20 年以上的每日数据历史。不仅如此，相关性还应二十年保持不变，收益也必须服从高斯过程。显然，这些方法依据的是不切实际的假设。

为便于理解这一问题，请参考图 11 中绘制的特征值分布。该图显示基于因子的财务相关性矩阵中，超过 90% 的特征值都服从马尔琴科 - 巴斯德分布。马尔琴科 - 巴斯德分布预测随机相关性矩阵中的特征值分布。换言之，几乎所有的特征值都与噪声有关——即使应用将回报结构化的因子模型后仍是如此。

图 11：在基于因子的相关性矩阵上计算的特征值分布



为纠正这一问题，Potter、Bouchaud 和 Laloux 建议去除相关性矩阵的噪声。³⁸ 操作流程包括使用一种机器学习算法，即核密度估计（KDE）算法，区分与噪声关联的特征值以及与信号关联的特征值。然后将与噪声关联的特征值全部替换为它们的平均值，从而使它们居中。图 12 绘制了去噪前后基于因子的相关性矩阵的特征值。收缩³⁹与去噪的区别在于，前者会求所有特征值的平均值，包括与信号关联的特征值，而后者仅求与噪声关联的特征值的平均值。一般而言，我们应优先使用去噪而非收缩，因为去噪有利于稳定协方差矩阵，同时又尽量减少了信号的丢失。

34 True Positive Technologies, LP 首席投资官、康奈尔大学教授、《Advances in Financial Machine Learning》（Wiley, 2018 年）和《Machine Learning for Financial Researchers》（剑桥大学出版社，2019 年即将出版）等书的作者。他是 AQR Capital Management 的负责人之一，也是该公司机器学习部的第一任负责人。此案例根据我们对他的采访以及他提供的资料编写。

35 H.M. 马科维茨，“Portfolio Selection”，《Journal of Finance》第 7 期（1952 年 3 月）：第 77-91 页，以及 F. 布莱克和 R. 利特曼，“Asset Allocation: Combining Investor Views with Market Equilibrium”，《Journal of Fixed Income》第 1 期（1991 年秋）：第 7-18 页。

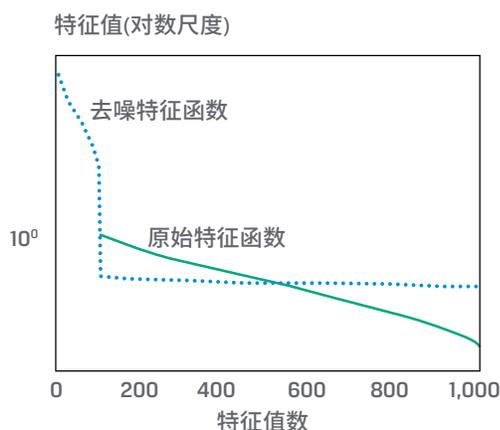
36 R. Michaud，《Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation》（英国牛津：牛津大学出版社，1998 年）。

37 V. DeMiguel, L. Garlappi 和 R. Uppal, “Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?” 《Review of Financial Studies》第 22 期（2009 年 5 月）：第 1915-53 页。

38 M. Potter, J.-P. Bouchaud 和 L. Laloux, “Financial Applications of Random Matrix Theory: Old Laces and New Pieces”, 《Acta Physica Polonica B 36》（2005 年 9 月）：第 2767-84 页。

39 例如参见 O. Ledoit 和 M. Wolf, “A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices”, 《Journal of Multivariate Analysis》第 88 期（2004 年 2 月）：第 365-411 页。

图 12：相关性矩阵的去噪



去噪可以保留相关性矩阵的轨迹，同时减少矩阵的条件数。得出的相关性矩阵将有更好的数值属性。我们可以通过下面的蒙特卡罗实验来证明这一点：(1) 使用基于因子的相关性矩阵，绘制一个随机经验相关性矩阵和均值矢量图；(2) 对经验相关性矩阵去噪；(3) 生成第 (2) 步的最大夏普比率投资组合；(4) 计算真实最优权重（从基于因子的相关性矩阵得出）和估计最优权重（从去噪后的经验相关性矩阵得出）之间的均方根误差（RMSE）。

表 1 列举了 1000 次蒙特卡罗实验的结果，显示去噪的效果远远好于收缩：去噪后的最大夏普比率投资组合的 RMSE，仅为未去噪的最大夏普比率投资组合的 0.04%。去噪本身就将 RMSE 降低了 94.44%，而使用 Ledoit-Wolf 收缩仅降低了 70.77%。虽然收缩在不使用去噪时有一些作用，但与去噪一起使用时没有效益，因为收缩会稀释噪音，但代价是也稀释了部分信号。

表 1：去噪和收缩组合的 RMSE 比较（最大夏普比率投资组合）

	未去噪	已去噪
未收缩	9.48E-01	5.27E-02
收缩	2.77E-01	5.17E-02

信号导致的投资组合失稳

凸优化解决方案的失稳还有第二个原因：信号导致的失稳。下面直观地解释了信号如何导致均方差优化失稳。当相关性矩阵为一个单位矩阵时，特征值函数是一条水平线，条件数为 1。在此理想场景之外，条件数会受到不规则相关性结构的影响。特别是在金融领域，证券子集表现出的证券间相关性大于其他投资范围时，该子集构成相关性矩阵内部的一个簇。簇会自然出现，层级关系的结果也是如此（例如，MSCI 行业分类的树形结构）。当 K 只证券形成一个簇时，它们会更严重倾斜于某个共同的特征向量，这意味着关联的特征值解释了方差更大的原因。但由于相关性矩阵的轨迹恰好为 N ，特征值只能在牺牲其他 $N-K$ 特征值的情况下增加，导致条件数大于 1。因此，簇内的相关性越高，条件数越多。这里要再提起，噪声导致的失稳来源于观察点太少。信号导致的失稳则来源于完全不同的原因，即相关性矩阵的聚类。

为解决这一问题，我们可以应用 López de Prado (2019 年起) 提出的嵌入式簇优化 (NCO) 程序：(1) 使用与 López de Prado 和 Lewis 所述类似的机器学习算法将相关性矩阵聚类⁴⁰；(2) 对每个簇分别应用优化算法；(3) 应用优化权重以将相关性矩阵展开为每个簇一行（和列）；(4) 对展开后的相关性矩阵应用优化算法。优化权重为第 (2) 步和第 (4) 步中权重的点积。这一程序可以提供稳健权重的原因，在于失稳的根源已经被控制在每个簇中，因此相关性矩阵的行为会较为理想。

40 M. López de Prado 和 M. Lewis, “Detection of False Investment Strategies Using Unsupervised Learning Methods”, 《Quantitative Finance》(2019 年即将发表)。可在 <https://ssrn.com/abstract=3167017> 查阅。

我们可以通过之前描述的蒙特卡罗实验来证明此方法的有效性。表 2 列举了 1000 次实验的结果。NCO 程序计算了最大夏普比率投资组合的马科维茨 RMSE 为 45.17%（即 RMSE 下降了 54.83%）。收缩和 NCO 组合使用可将最大夏普比率投资组合的 RMSE 减少 18.52%，优于收缩，但不如 NCO。同样，NCO 降低 RMSE 的幅度要大于马科维茨的模型，而收缩不增加价值。

去噪和 NCO，都可以显著降低估计最优投资组合的 RMSE

主要经验

马科维茨的投资组合优化框架在数学上是成立的，但其实际应用存在诸多问题。特别是由于噪声和信号的原因，金融协方差矩阵的条件数非常高。这些协方差矩阵的倒置会放大估算误差，导致解决方案失稳：修改观察点矩阵中的少数行可能导致完全不同的配置。即使配置估算无偏差，与这些失稳解决方案有关的方差毫不意外地造成大量交易成本，这可能吞噬这些策略的很大一部分盈利。

表 2：最大夏普比率投资组合的 RMSE

	马科维茨	NCO
原始	7.02E-02	3.17E-02
收缩	6.54E-02	5.72E-02

基于因子的相关性矩阵（例如 Barra 计算的那些矩阵）非常形象地解决了噪声导致的失稳问题，但没有解决信号导致的失稳问题。在本文中，我们讨论了两种方法，一种是去噪，另一种是 NCO，这两种方法都可以显著降低估计最优投资组合的 RMSE。从业人员应当熟悉这两种方法，运用它们来解决具体的问题。

11. 专题焦点：使用智能搜索采集和处理信息

供稿人：杨永智⁴¹

人工智能引擎是使用大量数据并借助机器学习、计算机视觉、图像识别、自然语言处理和光学字符识别技术来构建的。它们有可能取代初级分析师需要花很大一部分时间进行的许多重复性手工劳动。

随着互联网变得越来越庞大，数据也变得更加密集。当前搜索结果中的关键数据要么没有被注意到，要么重复存在于多个地方。人工智能引擎经过训练后可以理解并从难度最大的来源提取信息，包括社交媒体以及其他网站，此外还可处理本地可用的信息。

人工智能引擎可以支持对公告、研究报告、财经新闻和数据库的综合搜索。这种类型的搜索引擎超越了文本信息的范畴，使用计算机视觉来搜索文档中嵌入的图谱和表格。此外，它还会通过根据证券、行业、公司、个人和媒体建立实体关系模型，挖掘另类数据的价值。

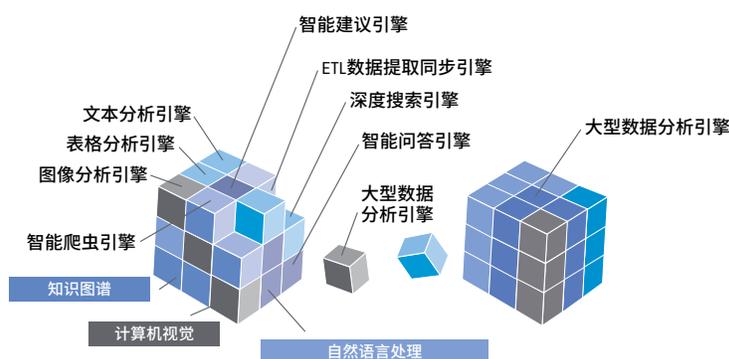
非结构化数据包括 PDF 文件、图像、照片、插图和表格、社交媒体贴文、博客、新闻等等格式的数据。人工智能引擎可让用户一次从成百上千个来源中搜索和提取关键数据。借助人工智能的强大力量，这些引擎还可跟踪到做出的任何更改并进行相应调整，最终可让分析师将重复工作自动化，更高效地利用自己的时间。

图 13 概括了我们的人工智能应用框架。

分析师的主要输出是研究报告。普通投资机构的分析师可能会使用多个门户来跟踪市场动态和公司信息。借助人工智能引擎（例如图 14 显示的引擎），可以提高下列领域的工作效率：

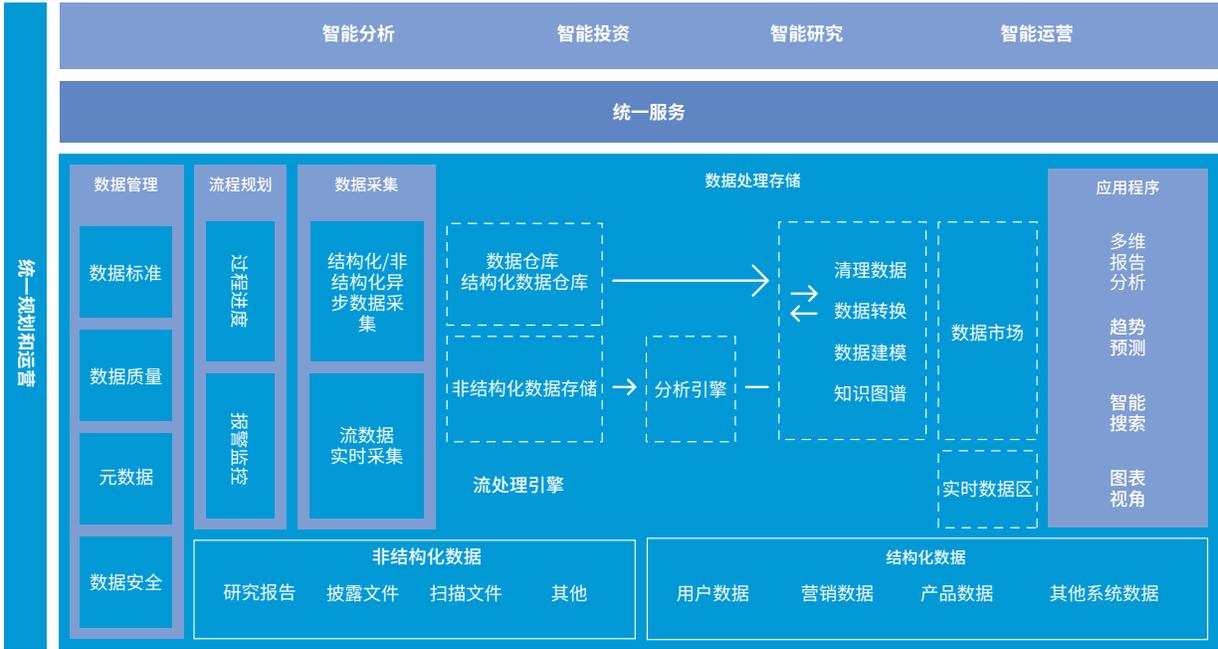
- 市场动态。分析师现在可以使用单个平台来搜索和提取多个来源的信息，并在几秒内将数据输入到模型中。
- 提醒。系统通常可以根据分析师的具体研究要求进行自定义设置，包括在检测到新信息时（例如政策变化）发出提醒。

图 13：阿博茨金融科技人工智能简介



⁴¹ 阿博茨科技创始人。此案例根据对他的采访以及阿博茨科技提供的资料编写。

图 14：分析师 .AI 框架



一些买方分析师依赖卖方的报告和公开渠道来进行研究。人工智能引擎可以让分析师进行更深入、更全面的搜索，从而扩大他们的研究范围。例如，分析师现在可以搜索报告中嵌入的信息，例如插图等。人工智能引擎还支持从报告的图表、表格中提取数据，极大地减少了分析师采集数据的时间，同时又提高了所采集数据的质量。以往分析师编制一份报告往往需要三至四周，现在最多需要一两周。此外，分析师还可以将更多时间用于实地走访和拓展投资想法。

例如，某位负责安防行业的分析师的工作不局限于与其关注公司及竞争对手管理层的会面，先于市场发现了行业趋势的变化。该分析师看到行业关注的焦点已经从安防摄像头的安装转向云技术，采购决策也日益向省、市、县集中。项目的规模和复杂性已经增加。这种趋势对拥有技术、大型项目管理经验和政府关系等竞争优势的新进入者有利。该分析师表示，他能够看到这些趋势，是因为他有时间进行更多调查研究，而不是将大部分时间花在收集信息上。

致谢

除各案例中列出的供稿人外，我们谨此感谢下列人员对本报告的无私支持：

Atit Amin, CFA

Ernie Chan

Simon Chao

Lisa Chen, CFA

Gordon Coughlan

Jan Deahl

Rebecca Fender, CFA

Francesca Guinane

Ariel Finno

鲍捷

姚军

Hinesh Kalia

Rob Langrick, CFA, CIPM

舒明, CFA

Jerry Pinto, CFA

Hugh Simon

Carol Ward

江旻

Kelly Ye, CFA

CFA INSTITUTE 工作人员

曹实, CFA, 金融分析研究高级总监
作者

Rhodri Preece, CFA, 金融分析研究高级主管
编辑

Gary Baker, CFA, EMEA 及金融分析研究董事总经理

ISBN 978-1-942713-78-4



9 781942 713784 >