

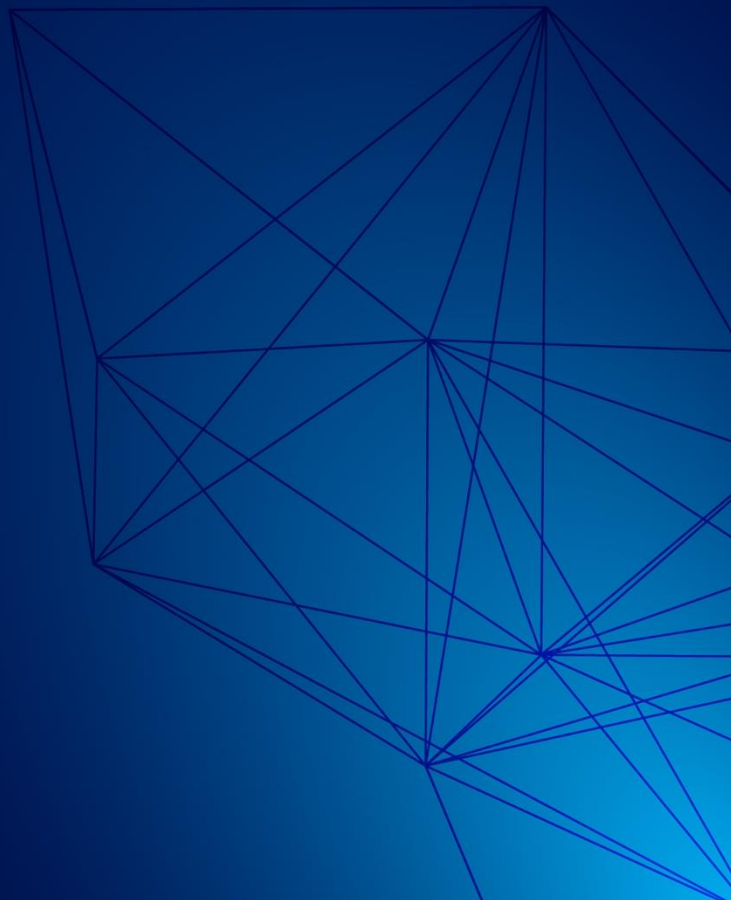


北大汇丰

PHBS FINANCIAL FRONTIER DIALOGUE

金融前沿对话

2019 年第 11 期 总第 22 期



PHBS HFRI
北京大学汇丰金融研究院

主办单位：北京大学汇丰金融研究院

院长：海闻

执行院长：巴曙松

秘书长：本力

编辑：都闻心（执行） 叶静 曹明明 鞠琳琳 方垚豪 朱伟豪

北京大学汇丰金融研究院简介

北京大学汇丰金融研究院 (The HSBC Financial Research Institute at Peking University, 缩写 HFRI) 成立于 2008 年 12 月, 研究院接受汇丰银行慈善基金会资助, 致力于加强国内外著名高校、金融研究机构、以及知名金融学者之间的交流, 构建开放的学术、政策交流平台, 旨在提高中国金融理论与政策的研究水平, 促进学术繁荣与发展, 加强与政府决策部门的联系与合作, 为政府决策提供参考意见, 为国际金融体制改革和中国金融业的发展做出贡献。

北京大学汇丰金融研究院院长为北京大学校务委员会副主任、北京大学汇丰商学院院长海闻教授, 执行院长为中国银行业协会首席经济学家、中国宏观经济学会副会长巴曙松教授。

AI 助力投顾甄选

主持：巴曙松（北京大学汇丰金融研究院执行院长、中国银行业协会首席经济学家、中国宏观经济学会副会长）

嘉宾：矫健（猫头鹰基金研究院投资总监）

一、问题的指出

（一）为什么能使用人工智能进行投顾甄选

第一，人工智能已经拥有更好的数据基础。这个数据不仅包含大量的详细交易数据，还包含大量的另类数据。第二，人工智能拥有足够的计算能力。人工智能所需要的大量计算在以前很难做到，但目前的硬件基础已经能够支撑其所需的计算，通过大量的训练、深度学习，训练自身神经网络。第三，人工智能已经开始出现在银行业、交易系统（尤其是高频交易），以及信用分析等领域。第四，很多海内外人工智能技术开始逐渐成熟、开放，这将为我们提供更好的应用条件。第五，在人工智能的应用以及在顶层的大类资产配置方面，已经有大量的研究基础。同时，在底层资产股票和债券方面，也已有较多人工智能方面的经验。

（二）为什么要用人工智能进行投顾甄选

中层的投顾选择非常困难，因为我们是在选择一个一个的基金经理，而人恰恰最难猜测。第一，人具有高度多变性，变化体现在人本身具有的复杂性，无法进行简单的概括。基金经理在进行投资时，有相当的艺术成分存在，即直觉成分，这是传统的量化方式难以进行捕捉的因素。同时，人具有很强的学习性，通过学习产生的变化难以用静态系统去捕捉。最后，投顾会进行风格漂移。第二，人是多维的。不同的人在不同的时空下所反映出来的能力不同，具有时空局限性。人在不同的市场阶段、宏观阶段所表现出来的能力有差异。基金经理的能力特征也是多维的，每一个基金经理都有能力边界，而每一个能力边界都约束了他所能够产生的投资效果，这个刻画是需要用比较强的人工智能学习体系去捕捉。第三，传统的研究方法，如定性的方法，已很难满足投顾甄选。投资经理如何从三千多支公募基金，两万多支私募基金中进行选择，海量的数据如何进行分析与比较，这是个难题。

AI助力投顾甄选时代到来

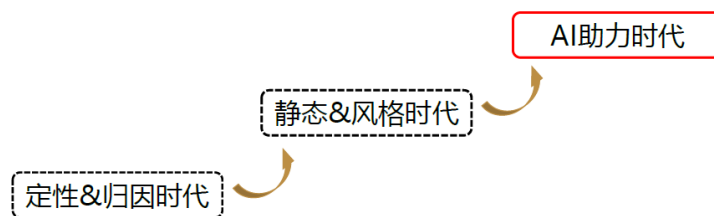


图 1

我们对整个投顾甄选的时代进行了划分。最早进行组合投资时，是定性和归因时代，简单的定性后进行投资。第二个时代是静态与风格时代。目前的投顾甄选基本上都停留在静态与风格时代。基金经理会画九宫格或者更多格子，对历史情况进行分析，并得出投资的内容。我们认为这个阶段也很难满足目前投顾甄选的需求。人工智能时代已经到来，人工智能可以助力投顾甄选。

二、海外的经验借鉴

从海外研究文献来看，美国和欧洲的研究更多集中在基础和大类资产上。在对组合经理进行选择，投顾甄选的环节做的不足。而日本养老金，即 GPIF 的做法非常值得我们学习。

（一）虚拟基金经理的训练与发现

海外的借鉴：虚拟基金经理-训练集与发现（GPIF）

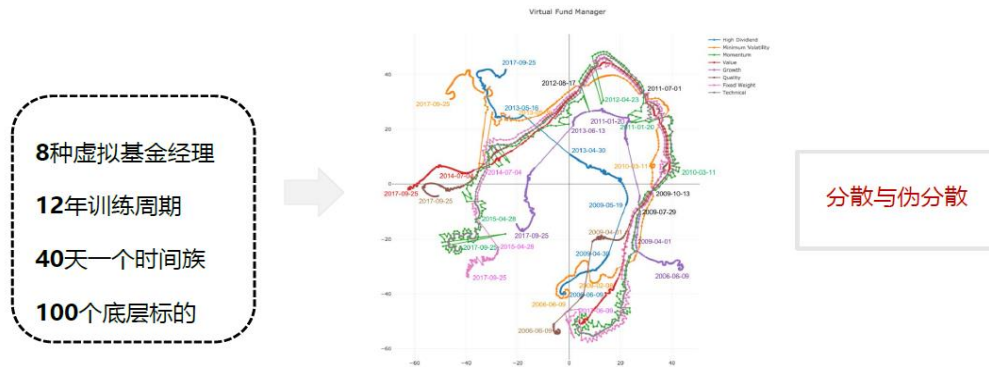


图 2

GPIF 进行了较为深入的研究，把风格大类分成如下几类，分别是高分红、低波动率、动量、价值、成长、质量、固定权重以及技术。他们把这八类分别称作八种虚拟基金经理，从 2005 年到 2017 年，以

十二年的时间作为训练周期，以每 40 天为时间族，100 个底层标的，19 个因子作为研究对象。以此训练这八种虚拟基金经理，观察将其放入一个训练集时会产生何种样效果。

通过图 2 中间的多维平面化展示图可以看出两点结论。第一，以八个大类风格去确定的八种虚拟基金经理，训练出来的神经网络具有很大价值。这样的神经网络可以进一步分析真正的基金经理。第二，从这八个大类所做出的研究结果可以看出，静态和风格划分的时代，缺失了很多信息。缺失的信息主要体现在清晰分类下，具有独立风格的虚拟基金经理，却在相当大程度上无法取得分散的效果。

（二）神经网络—风格探测矩阵

海外的借鉴：神经网络-风格探测矩阵（GPIF）



图 3

图 3 演示了 GPIF 如何进行风格探测。将八种虚拟基金经理的八种固定特征作为一个输入源，放入神经网络中，在不同的时间进行训练，训练得出风格探测器。过去与现在的数据里产生了较多的数据集。

该数据集是体现在具体的历史情境中，不同的情境发生了不同的事情。在情境中，每一个基金经理，做出的交易记录都放入风格探测中，最后它将输出“投资风格”。投资风格是多维的，基金经理在不同的维度展现出多大程度的复杂性，是历史的结果，且这个历史的结果，是一个概率矩阵。

（三）实例——稳定与漂移

海外的借鉴：实例-稳定与漂移 (GPIF)

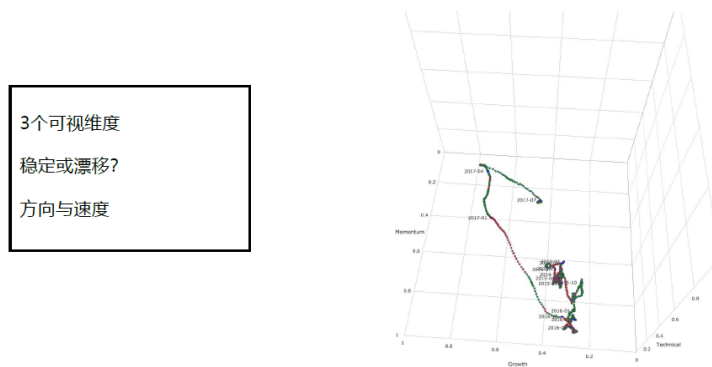


图 4

使用风格探测器，可以帮助我们查看具体的基金经理，在动量特征、成长特征、技术特征三个维度会如何表现，我们把它称作三个重要的可视维度。由图 4 可见，在 12 年测试期中的 10 年左右时间，基本上围绕在较小的区域内稳定地进行漂移。技术层面的数据会有小范围浮动，但成长维度在很小区域内移动，因此风格总体上稳定。但是从 2016 年的某一个阶段到 2017 年 7 月份，其风格发生了非常明显的漂移。漂移过程有快有慢，它以不同的方向和速度进行漂移。这告诉我们若想分析一个基金经理，用神经网络的体系才能够真正捕捉到他

的特点。通过研究产生的概率极预测了可能产生的漂移，这对于组合经理的意义非常巨大。

（四）再训练和行为预测系统

海外的借鉴：再训练和行为预测系统（GPIF）



图 5

如果预测的交易行为未发生，可以如图 5 中所示进行神经网络再训练。将大量数据放入外部情境中检测，检测好后，放入体系中进行训练，训练结束后出现“行为预测器”。该“行为预测器”是基于每一个指定的投顾，预测在外部情境下可能产生何种交易行为，以及产生某种交易行为的概率。这个外部情境就是组合基金经理对未来的判断。通过 AI 训练得出行为预测器，从而产生概率集，并预测未来可能做出的交易。这个系统对任何一个投顾甄选过程都产生了巨大作用。

（五）GPIF 的启示：AI 可以助力投顾甄选

GPIF 的研究给我们的启示是：AI 已经可以助力投顾甄选，主要表现在如下四方面。

1. 更加科学：用动态过程去替代静态事物，用预测辅助判断。

2. 更多维度：可以通过更多角度进行训练、分析，甚至可以包括大量的非交易信息，如信用、情绪等六类数据。这些数据的加入对于分析人的可变性、不可揣测性，将有极大好处。

3. 更易跟踪：经过更加精准、多维的风格识别，更准确地进行风格划分。

4. 更易预测：基于场景化的行为预测以及组合管理。

三、当前中国国内的探索

（一）原则

1、坚持金融自洽性：以行为金融理解为边界，形成闭环。

2、坚持实用性原则：中国与国外有很大区别，无论是政治周期、投顾经验、能力边界、管理的时限以及考核期等，都需要考虑入内。

（二）问题和难题

1、小样本问题。日本养老金有大量数据作为基础，但是中国没有如此多数据。因此必须用处理小样本的技术去处理小样本问题。

2、经济学模型与学习模型结合问题。国内外都面临着该问题。学习模型是经验模型，而经济学模型是逻辑模型，逻辑模型与经验模型在相当大程度上不能完全结合。因此，在处理人工智能所输出结果时，需要考虑如何衔接与解决矛盾。另外，中国的经济学，往往用国外的逻辑无法解释。因为中国经济学受到多方因素影响，包括政策干预、企业行为、政府行为等。此类差异使得中国传统经济学逻辑会有一些问题。当然经济学模型跟学习模型结合的问题，在中国未见得比

海外更严重。原因在于我们可以忽略一些海外成熟的经济学模型，更多使用人工智能的学习模型去代表。用人工智能所能解释的内容，可能比生搬硬套海外理论更切合实际。

3、不明确的边界。目前人工智能解决的问题，绝大部分问题都是有边界的。比如国际象棋、围棋、视觉处理、图像处理、声音处理、视频处理。但是做组合投资，分析金融资产，以及分析管理金融资产的投资人时，边界将非常难以确定。

4、投资的艺术性如何与硬科学相衔接。

人工智能助力投顾甄选，主要就从以下几方面出发：（1）通过人工智能取得动态分析投顾，多维度、精确地分析投顾的风格，将其放在较好的神经网络里进行刻画。（2）在具体的情境下，在投顾进行情境输出时，能够做好概率输出，具备一定的预见性。（3）加入人工干预，做成半闭环的形态，在此期间可输入一定东西，提高系统正确度。

四、中国的实践

猫头鹰已经完成了对基金（经理）69-398 个不同的本土化指标的算法建立、验证与入库工作。平均每个基金（经理）163 个本土化指标（每单位时间）。这些指标刻画出了基金经理行业维度视角，例如行业维度分析、smartbeta 分析、适应性分析等。

年份	轮动指数	价值强势	成长强势	配置方案	code	name	fund_co	code	name	fund_co
2006			FALSE					30040415	王亚伟	NA
2007	0.763	FALSE	FALSE	重成长	30040205	邓晓峰	NA	30040209	董承非	兴全基金
2008	-0.127	FALSE	FALSE	偏价值	30037138	程世杰	NA	30040529	郑煜	华夏基金
2009	-0.133	FALSE	FALSE	偏价值	30038788	杨毅平	NA	30038184	宋小龙	NA
2010	-0.787	FALSE	FALSE	重价值	30040296	林飞	NA	30040539	周蔚文	中欧基金
2011	-0.078	FALSE	FALSE	偏价值	30050178	胡涛	嘉实基金	30045196	王卫东	NA
2012	0.189	FALSE	FALSE	偏成长	30053914	杨靖	NA	30038823	于鑫	NA
2013	-0.240	FALSE	TRUE	重成长	30062004	朱颖	NA	30044188	刘红辉	诺安基金
2014	-0.286	FALSE	FALSE	重价值	30057348	陈明星	NA	30040517	张勇	NA
2015	0.105	FALSE	FALSE	偏成长	30056013	温胜普	NA	30040413	王晓明	NA
2016	-0.531	FALSE	FALSE	重价值	30042608	张惠萍	NA			
2017	0.065	FALSE	FALSE	偏成长						
2018	0.687	FALSE	FALSE	重成长						
2019	-0.129	FALSE	FALSE	偏价值						

图 6

通过上述系统，经过大量历史数据归演，产生结果如上。在输入情境时，根据某一些固定的指标变化，去选择第二年的情境可能如何表现，并根据推出的情境进行配置。旁边这两张图是 2012 年输入特定情境时，我们需要筛选的情境其所具有的价值性与成长性。例如邓晓峰、陈世杰、董鹏飞、宋小龙，然后把其相结合后去匹配 2012 年的数据。每年进行推演、预测，并得到图 7 的结果。

猫头鹰ALPHA筛选器表现

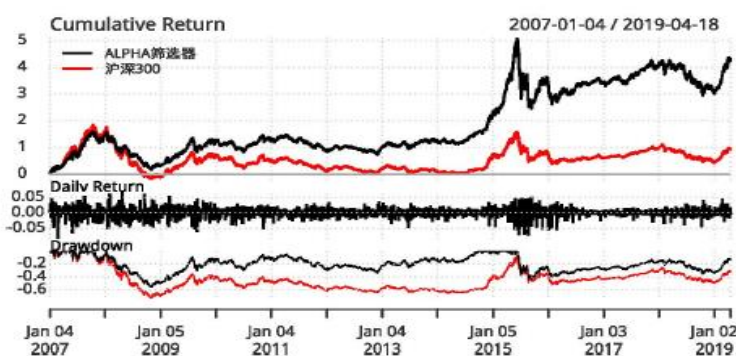


图 7

从图 7 中数据表现来看，学习系统在不停地积累数据的过程中，从波动性与收益率上都反映出了一定优于指数的特征。这为投顾甄选者即 FOF 基金经理、银行理财子公司、券商资管及保险公司都提供一

定启示。

启示主要包含三方面。第一，多维刻画，避免盲人摸象。基金经理会动态地在不同情境下面做出反应。同时他们有很多且精细的维度去刻画。第二，避免刻舟求剑，我们要预测变化，而不是等待变化。第三，全面刻画，利于教学相长。公私募的管理人成长过程自下而上，从行业或公司研究到基金处理，再担任基金经理。其缺点是缺乏组合思维跟配置能力。此外，作为委托方的银行、保险等专家，他们具备宏观的视角，但是缺乏微观的支持与投资落地。通过 AI 助力投顾甄选，恰好可以达到教学相长的目的。通过与投资经理沟通，能够得到优化资产配置的能力，理解底层资产的能力，分析行业和个股的能力。此外，通过横向比较投资经理，获得更全面的市场画像，得到更深刻的市场结构理解，识别并学习真正的优秀群体。

五、问答环节

Q1: 您这个研究重点是关注预测基金经理的行为，而不是他们的业绩表现是吗？使用者需要先预测未来的场景吗？

A1: 业绩表现是一个结果。现在静态的分析往往难以取得很好的效果，正是因为没有精确对基金经理进行动态多维的刻画。这是 AI 助力甄选最不成熟的部分即投顾行为的预测。这个技术本身很前沿，是一个重要的研究方向。在给定场景的概率情况下进行行为预测。如果投顾对未来没有清晰的判断，也可以是一个等概率的场景判断。如果有了预判，就可以加入这个预判。这样的预判将帮助提升效果，但

本身不是必须的。

Q2: 行为具体是指什么，例如最空选哪类股票或选哪些行业。

A2: 这个行为本身是交易行为。既可以是多空，也可以是交易不同的底层资产的过程。既可以包括行业因素也可以包括个股因素。具体是在这个时空背景下，最有可能操作的行为。这块是我们研究中正在展开的一些工作。

Q3: 您刚才谈到要去预测变化的发生，而不只是等待。基金经理在他的从业年限里是会学习的，例如上一次在某种情境中表现不好，下一次出现同样情境时可能会改进甚至超过平均水平。请问这类关于基金经理学习成长的概率 AI 可以反映吗？

A3: 所谓的训练集和神经网络实际上是一个学习机制。输入和输出是历史的素材，但是历史的材料可以看到很多具体投顾的关于学习能力的内容，足够的数据就可以对未来的能力增长进行一定的预判，但这只是一个模糊的判断。更多是根据已有的判断将学习能力进行预判。而不是预判他还能学到什么。

Q4: 这个研究对于基金的数据要求比较高，如何满足在实践中能够及时发现判断失误？

A4: 如果是 MOM 模式，就没有数据问题，只需考虑时间长短问题。如果是 FOF 模式则需要运用持仓穿透技术进行估计。但对于学习器，若数据不足，不准确，风格分类器和预判器也只能跟随数据的精度而定。但运用了一定技术可以解决一部分问题，但无法解决精准问题。

Q5: 加强对投顾的动态了解是否可能带来 FOF、MOM 频繁更换投顾，增加的交易成本对冲掉相关的好处？

A5: 要约束条件，更换本身也有成本。更多的是事前匹配更合理、刻画准确的投顾，并进行事后管理。本质是所需要的特性若不再能提供就需要调整。把成本和考核周期作为约束条件。

本文根据北京大学汇丰金融研究院执行院长巴曙松教授发起并主持的“全球市场与中国连线”第二百九十四期(2019年5月24日)内容整理而成，特邀嘉宾为猫头鹰基金研究院投资总监矫健先生。

矫健先生拥有12年证券公司和基金公司投资和研究经验，曾先后就职于申银万国证券研究所、交银施罗德基金和博道投资等机构，现任猫头鹰基金研究院投资总监。猫头鹰基金研究院致力于从量化和定性双视角理解基金经理，做AI（人工智能）化的“中国的晨星+中国的罗素”，争取成为最懂基金经理的独立第三方研究机构。

【免责声明】

“全球市场与中国连线”为中国与全球市场间内部专业高端金融交流平台。本期报告由巴曙松教授和王志峰博士共同整理，未经嘉宾本人审阅，文中观点仅代表嘉宾个人观点，不代表任何机构的意见，也不构成投资建议。

本文版权为“全球市场与中国连线”会议秘书处所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复印、发表或引用本文的任何部分。



PHBS HFRI
北京大学汇丰金融研究院

