

机器学习驱动的农业经济学研究范式革新

陈梓萌 沈宇辉 蒋卓人 龚斌磊

摘要: 农业农村的快速发展与农业经济研究问题复杂性的不断提升,使得传统研究范式在应对高维结构与复杂模型设定方面的局限日益凸显。机器学习凭借其深度数据挖掘、非线性建模与预测分析等优势,能够为相关研究提供有力的技术补充与方法拓展。本文按照开展研究的逻辑顺序,从数据收集处理、模型构建等环节系统梳理和比较机器学习与传统农业经济学研究方法,并探讨二者在关键研究场景和重点研究领域中的融合可行路径与协同机制。通过对典型应用场景与重点领域的考察,本文提出机器学习嵌入农业经济学研究的协同机制,并凝练出机器学习驱动下预测与评估协同、跨模态多源信息融合、多任务系统集成的三大转型趋势。本文探讨了机器学习对农业经济学研究范式革新的重要价值,并系统梳理农业经济学研究方法的发展脉络,可为推动农业经济学研究范式创新、加快构建中国哲学社会科学自主知识体系提供参考。

关键词: 机器学习 农业经济 研究范式

中图分类号: F3; C3; F224-3 **文献标识码:** A

一、引言

农业经济学(以下简称“农经”)兼具深刻的理论复杂性与重要的现实意义,呈现出显著的学科交叉特征(钱文荣和郑淋议,2025)。其核心任务在于既要解析农业生产中自然、经济、社会交织的多元矛盾,又要回应国家粮食安全、农村发展和农民

[资助项目] 国家自然科学基金国际重点项目“中国农业绿色转型与高质量发展的路径与战略研究”(编号:72161147001);国家自然科学基金面上项目“农业技术扩散与生产率收敛:理论与实证研究”(编号:72173114);国家自然科学基金青年项目“基于多智能体协作的政策感知与反馈机制研究”(编号:72574198)。

[作者信息] 陈梓萌,浙江大学公共管理学院,电子邮箱:chenzimeng@zju.edu.cn;沈宇辉,浙江大学公共管理学院,电子邮箱:yuhuishen@zju.edu.cn;蒋卓人,浙江大学公共管理学院,电子邮箱:jiangzhuoren@zju.edu.cn;龚斌磊(通讯作者),浙江大学中国农村发展研究院、浙江大学公共管理学院,电子邮箱:gongbinlei@zju.edu.cn。

生计等重大战略命题（罗必良，2020）。在新发展阶段，“三农”问题的系统性、动态性和非线性等复杂特征日益凸显，传统以线性建模与结构化数据为基础的研究范式逐步显现出局限性。换言之，在各类跨界技术深刻影响农业农村的背景下，农经研究本身也需要引入跨界技术，以适应研究对象与研究问题的变化。与此同时，机器学习作为数据驱动的方法体系（Mullainathan and Obermeyer, 2017; Athey and Imbens, 2019），在复杂系统建模、高维特征提取与非线性关系识别等方面展现出独特优势，逐渐成为农经研究范式革新的重要驱动力。需要强调的是，农经研究范式的这一革新并不是对传统农经理论和实证方法的颠覆，而是驱动农经研究方法与研究范式的持续更新与融合，助推农林经济管理学科（以下简称“农经学科”）跨越式发展。在此背景下，探索机器学习方法与传统农经研究分析工具的融合路径，对于拓宽研究边界、深化问题认知、优化政策设计等具有重要的理论价值与实践意义。

一方面，农经学科始终强调对现实问题的关注与回应，其研究目标不仅在于解释经济现象，更关注未来应当如何做。传统计量方法以因果识别为核心，在应对情境模拟和复杂系统预测方面相对不足（郭峰和陶旭辉，2023）。机器学习则强调预测性能与模型表现，在政策模拟、农产品价格预测等应用中，展现出巨大应用潜力（Imbens, 2015; Zheng et al., 2023）。另一方面，农经研究的多学科交叉特征为机器学习方法的融入提供了广阔空间与独特机遇（Wuepper et al., 2025）。农经研究需要深入解析自然、经济、社会多维系统，这决定了其研究过程面临着数据维度高、信息跨模态以及目标多元化的挑战。面对传统方法难以兼顾的瓶颈，学科发展亟须研究范式的升级。而机器学习在处理高维非结构化数据、动态预测、融合跨模态信息和多目标协同优化方面的比较优势，能够为破解上述难题提供可行路径。

当前，已有部分研究将机器学习应用于农经研究的不同领域，包括农业生产（Gulati et al., 2025）、农产品价格（Paolo et al., 2024）、食品消费（Tao et al., 2024）等热点主题。这些研究不仅提升了预测效果，也为政策制定和治理实践提供了有效支持，展示出机器学习在农经学科不同研究领域中的广泛应用潜力。但与此同时，机器学习在实际应用中仍面临数据质量不稳定、模型可解释性有限等问题（Athey, 2017; Northcutt et al., 2021），这也是未来农经研究需要进一步优化与拓展的方向。尽管已有初步探索，中国农经研究中机器学习的应用仍处于起步阶段，相关研究成果的数量和影响力仍然有限。究其原因，主要在于当前学术界对机器学习方法的认知尚不系统，缺乏对农经研究特性和机器学习之间契合点的深度梳理，且容易将机器学习方法与传统农经研究范式视作对立关系。事实上，机器学习对于农经研究的意义，不仅仅是工具箱的扩充，更是研究范式的革新，它促使研究从传统计量分析的假设驱动转向数据驱动，并在突破预设函数约束的同时，推动跨模态多源信息融合与多目标协同评估。

鉴于此，本文围绕机器学习与农经研究的融合展开系统分析，从数据基础重构、模型方法演进、典型应用场景与重点研究领域等方面，探讨机器学习驱动农经研究范式革新的内在逻辑与实现路径。相较于现有文献，本文的主要贡献体现在两个方面：第一，本文厘清了机器学习在农经研究中的独特适配性与应用逻辑。有别于现有文献多聚焦于一般应用经济学领域的讨论，本文立足于农经学科，从研究对象的动态异质性、研究目标的实践导向性以及学科属性的交叉融合三个维度，深度剖析了引入机器学习的必要性。第二，本文系统梳理农经研究横跨“自然、经济、社会”三维系统的特性，结合机器学习的特征，探讨其对农经研究范式的深层驱动，并分析其在应用实践中的潜力。

为方便读者阅读与理解，本文按照开展研究的逻辑顺序，从以下五方面展开论述：第一，在数据获取与处理方面，分析机器学习在非结构化与多模态数据整合中的能力及其对农经研究数据体系的重构作用；第二，在理论建模方面，比较机器学习方法与传统计量方法在模型设定、变量选择与算法设计方面的异同与互补性；第三，在研究领域方面，示例性聚焦气候变化与资源环境、农村劳动力转移与农业转移人口市民化、农业生产与农业技术进步三个领域，探讨与展望机器学习的应用价值；第四，在研究场景方面，重点探讨机器学习方法在政策效应评估与动态预测、跨模态信息融合与建模、多任务学习与综合评估三类典型研究场景中的比较优势与嵌入逻辑；第五，在研究前瞻方面，结合国内外最新进展，总结在机器学习方法驱动下农经研究的三个转型趋势，旨在为推动农经学科方法创新与交叉融合提供理论借鉴与实践参考。

二、机器学习驱动的数据基础与获取方式变革

农经学科的研究对象天然横跨自然、社会和经济多重维度。正因如此，农经研究具有显著的学科交叉属性，需在立足自身学科体系的同时，深入理解自然再生产和经济再生产的耦合机制。农业经济活动受自然条件影响更深、时空异质性更强、产出过程更具非线性，因此，其研究对象的复杂度远超其他学科。随着大数据时代的到来，“三农”领域的的数据呈指数级增长，海量多源数据正在以前所未有的规模和复杂度涌入研究者的视野，面对如此复杂的高维信息，依赖结构化数据与简化模型的传统研究框架日益捉襟见肘。在此背景下，如何高效获取、整合并利用海量数据，已成为制约农经研究进一步深化的关键瓶颈。机器学习方法正是在这样的现实条件下被引入农经学科，机器学习的兴起是农经研究对象复杂化后的方法论响应，也是提升学科解释力与指导力的内在需求。

（一）数据来源的演变：从结构化调研到多模态观测

目前农经传统研究的数据主要来源于问卷调查、官方数据、田野观察等（王芳

等, 2020), 采集成本高、样本代表性较弱、隐私要求严格。具体而言, 常见的数据类型主要包括调查数据、行政统计数据 and 访谈数据等。总体而言, 当下农经领域的实证研究仍高度依赖结构化数据, 这类数据往往有明确的格式和指标, 经简单处理后能够直接用于定量分析 (刘景江等, 2023)。然而, 该传统研究范式也存在一些难以忽视的限制: 一方面, 研究数据更新周期长、空间分辨率低, 难以追踪农业生产和市场变动的实时特征; 另一方面, 在面对图像、文本、音频等非结构化数据时, 往往力有未逮, 缺乏处理能力和方法支持。这些限制使得传统方法在处理高维异构数据、捕捉非线性互动关系与刻画时空动态模式方面表现不足, 也为新方法特别是机器学习的应用留出了广阔空间 (方帅等, 2025)。

相比之下, 机器学习能够处理大规模、异质性强且结构复杂的数据, 其数据来源更加多元化, 涵盖卫星遥感影像、无人机航拍照片、物联网传感器记录等多种类型。图像数据是当前应用最广的一类非结构化数据资源, 包括地表覆盖、作物识别、气候灾害评估等关键任务中广泛使用的多时相遥感卫星影像、农业地块高分辨率航拍图像等; 文本数据则包括农业政策文件、地方政务信息、新闻报道与社交媒体评论等, 这些文本资源经过采集与挖掘, 能够反映政策取向、社会关注点与消费者行为 (Ludwig et al., 2024)。此外, 还有越来越多农机设备、智能气象站、移动终端等产生的时间序列数据与传感器监测数据, 这些非结构化数据虽然具有潜力, 但在实际研究中存在格式不统一、噪声干扰大等挑战。例如, 遥感影像需要进行几何校正才能用于建模, 文本数据需经过分词等预处理环节才能生成结构化变量。

(二) 非结构化数据的转化：机器学习驱动的特征提取

随着数据来源日益多样化, 尤其是图像、文本、音频等非结构化数据大量涌现, 亟须新的方法将这些原始数据转化为可用于建模的经济学变量。机器学习为非结构化数据的结构化转化提供了新的解决方案。通过卷积神经网络、自然语言处理模型、图神经网络等工具, 研究者可以从遥感图像中识别土地利用类型 (Jia et al., 2019), 从文本中提取政策语义信息 (Tao et al., 2024), 从物联网设备数据中捕捉行为模式, 进而生成可用于经济建模的变量。这种基于数据驱动的特征提取方式, 极大拓展了传统农经研究的变量维度与解释能力, 为识别复杂经济机制与提升预测性能提供了坚实支撑 (洪永森和汪寿阳, 2021; 刘景江等, 2023)。

在图像数据方面, 机器学习方法同样为农经研究带来一定突破。以往农经学者多依赖遥感领域已解译的土地利用或作物类型数据。这些数据扩大了时空覆盖范围, 但主要服务于地球科学关注的宏观变化, 对农经学科重点关注的地块经营、农户福利等微观议题支撑有限。近年来, 随着深度学习、公共卫星影像库和云端算力的普及, 越来越多的农经学者将遥感影像纳入研究之中。通过机器学习的方法, 研究者可以从复

杂的图像数据中生成感兴趣的研究变量，为许多研究话题提供新的可能。例如，Huang et al. (2021) 使用机器学习方法分割卫星图像中肯尼亚农村地区住房的轮廓并分析其动态变化，进而评估了肯尼亚农村反贫困计划对农户福利水平的影响。

在文本数据方面，研究者关注的核心常常是文本主题、文本相似度等。在农业政策分析中，研究者可以通过网络爬虫获取海量政策文件和新闻文本数据。机器学习模型据此生成关于政策影响的定量指标，例如Tao et al. (2024) 使用无监督机器学习方法，定量识别各省新闻媒体报道中涉及食品安全主题的信息数量，进一步将其纳入家庭消费的研究中，研究新闻媒体有关食品安全的报道强度对消费者外出就餐的影响。在农户行为研究领域，机器学习方法能够从电信记录、社交媒体使用数据等大样本行为数据中，提取更丰富的农户决策特征（王芳等，2020；Storm et al., 2024）。为更清晰地展示不同类型数据的特点，本文总结了当前农经研究数据类型的主要特征及其处理方式。具体如表1所示。

表1 农业经济学研究的数据类型、结构特征与处理方式对比

数据类型	典型来源	是否结构化	常用处理技术
问卷或统计数据	调查数据、农业普查数据、统计年鉴	是	线性回归、双重差分法
图像数据	遥感影像、多时相卫星图、无人机图像	否	卷积神经网络（CNN）、编码—解码式卷积网络（U-Net）
文本数据	政策文件、政府公报、新闻、社交媒体	否	双向编码器表示转换器（BERT）、大语言模型（LLM）
时序监测数据	气象数据、传感器数据、农机记录	否	循环神经网络（RNN）、长短期记忆网络（LSTM）
音频数据	畜禽咳嗽声、农机噪音、环境监测音频	否	频谱图+CNN、自监督音频建模

需要注意的是，非结构化数据的可获取性并不代表其天然具有经济学价值，其与实际经济行为之间的映射关系仍需理论检验与现实验证。机器学习驱动下的变量构建，应以明确的研究问题与机制识别为前提，避免“为技术而技术”的工具导向倾向，防止模型陷入脱离理论的黑箱化操作，从而削弱研究的解释力与政策参考价值。

（三）数据质量挑战与标注机制反思

机器学习拓宽了数据获取边界，但其依靠数据驱动而非先验设定的特性，使得模型对数据质量尤为敏感。如果原始数据存在测量误差或系统性偏差，算法将这些噪声纳入拟合函数，进而会导致预测产生系统偏误（郭峰和陶旭辉，2023）。特别是对监督学习而言，高质量的数据标签是保证模型有效的先决条件——标签质量低会导致偏误，削弱模型识别真实结构的能力；而样本量不足则会导致机器学习方法的估计方差增大，削弱外推能力。若无法通过数据清洗、再抽样等方式有效纠正偏差，即便机器学习模

型具备高度灵活性，也只是高精度地拟合一个与现实完全背离的函数形式。

因此，在机器学习应用情境中，标签质量构成关键制约因素。数据标注已成为机器学习应用于农经研究的核心问题之一，也是在评估传统研究方法与新方法效能时需深入权衡的关键维度。传统范式强调严格采样与精确定义，机器学习则需引入新的质量保障机制，避免“垃圾输入，垃圾输出”。未来，需要在农经领域建立更加完善的数据标注与验证机制。例如，在作物识别、病虫害监测等需要人工标注的任务上，引入专家标注与志愿者标注相结合的模式；在农户调查中，应用移动设备实时采集和交叉验证关键信息，减少人为误差。同时，随着技术进步，无监督或自监督的方法可能降低对精确标签的需求，允许模型从未标注的数据中学习并增强泛化能力，然而这类方法的具体效果仍需进一步评估。未来，农经领域亟须构建标准化、可追溯的数据管理机制，为跨学科建模提供坚实基础。尤其在非结构化数据广泛应用的背景下，如何协调多样数据与经济学含义的对应关系，避免数据复杂性与经济学逻辑脱节，将成为评估机器学习方法有效性的重要标准。

三、机器学习驱动模型构建与分析方法演进

随着农经研究主题的拓展与研究数据日益高维、多源，传统计量经济方法在处理非线性关系、高度异质性样本及复杂交互项时显现出一定的局限性（于晓华和刘爽，2024）。因此，机器学习逐渐成为补充分析框架的重要工具（Athey and Imbens, 2019; Mullally et al., 2021）。本文从农经传统研究范式出发，聚焦传统计量分析和机器学习方法在模型构建与算法设计方面的差异展开系统梳理。两类方法的核心差异如表2所示。

表2 传统计量分析方法与机器学习方法在农业经济学研究中的对比

比较维度		传统计量方法	机器学习方法
研究逻辑	研究出发点	以假设驱动为核心	以数据驱动为核心
	研究对象	结构化数据	结合非结构化数据
	核心目标	事后因果解释	事前预测与异质性挖掘
研究设计	变量选择	理论导向与逐步回归	自动筛选变量
	函数设定	预设函数形式	无须预设函数形式
	参数估计	无偏估计	偏差—方差权衡
模型特性	可解释性	强：参数具有明确含义	弱：部分模型为黑箱
	高维适应能力	弱：受自由度约束	强：适配高维数据
	非线性处理能力	弱：依赖先验函数设定	强：擅长刻画非线性关系

（一）理论驱动的建模方法：以线性回归为代表的传统范式

计量分析是农经领域广泛采用的研究方法，主要包括线性回归模型、离散选择模

型和匹配方法。计量分析方法大多是依托明确的统计模型，通过参数化的方式刻画变量之间的关系。以计量经济学中广泛采用的线性回归模型为例，线性回归模型是计量分析中最基础且直观的实证工具，其基本形式可以表示为：

$$y_i = \hat{\beta}_0 + \sum_{k=1}^K \hat{\beta}_k x_{ik} + \varepsilon_i \quad (1)$$

(1) 式中， y_i 为因变量， x_{ik} 为第 k 个解释变量， β_k 为待估参数， ε_i 为扰动项。为保证参数估计具有一致性、无偏性和有效性，需要保证模型符合一系列经典假设：线性、随机抽样、无完全共线性、零条件均值以及同方差性。在保证满足假设的前提下，通过最小化误差平方和的方式寻找最佳参数，即用普通最小二乘法（OLS）拟合模型。具体公式如下：

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K} \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{k=1}^K \hat{\beta}_k x_{ik} \right)^2 \quad (2)$$

线性回归模型作为基础计量工具具有良好的可解释性，且其估计结果容易被验证（Wooldridge, 2010；仇焕广等，2018）。但是，当面对变量数量庞大、因果关系复杂的数据集，以线性回归模型为代表的计量分析方法可能会面临以下问题：第一，线性回归等计量模型要求研究者在建模前预设函数形式，这一先验设定容易产生模型设定偏误，导致遗漏关键变量，从而使估计结果有偏。第二，经典线性模型在建模时通常假设误差项满足正态分布、同方差和独立性等一系列统计条件，而自然条件下的高维数据往往无法完全满足这一系列条件。机器学习方法提出了以数据驱动为核心的建模分析策略，并通过灵活的建模方式和算法设计（Athey and Imbens, 2019），为研究者提供了重要的补充性分析工具。

（二）高维情境下的变量选择：正则化估计方法

传统线性回归模型的局限性之一在于参数估计的方差相对较大，也就是估计系数的上下波动较为剧烈（王芳等，2020）。在高维情境下，受自由度限制，普通最小二乘法求解所依赖的协方差矩阵无法求逆，导致模型无法正常求解。此外，高维情境下海量变量容易引发多重共线性问题，数据扰动对估计结果的影响被放大，一定程度上削弱模型的泛化能力。为应对这种情况，机器学习在目标函数的最小化过程中，通过加入估计系数的绝对值来抑制系数大小，从底层架构上突破自由度与矩阵求逆的约束。以最小绝对收缩与选择算子（LASSO）为代表的正则化方法，为传统计量分析方法在高维数据环境中的应用提供了重要拓展（Belloni et al., 2014）。其目标函数如下：

$$\hat{\beta} = \arg \min_{\hat{\beta}} \left\{ \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^K \hat{\beta}_j x_{ij} \right)^2 + \lambda \sum_{j=1}^K |\hat{\beta}_j| \right\} \quad (3)$$

(3) 式中, $\lambda \sum_{j=1}^K |\hat{\beta}_j|$ 为 L1 范数惩罚项, 当 $\lambda = 0$ 时, LASSO 等价于 OLS。随着 λ 增大, 部分 $\hat{\beta}_j$ 会被压缩, 从而实现变量选择。与 OLS 对变量的选择高度依赖研究者主观判断不同, 机器学习的 LASSO 方法通过引入惩罚机制, 利用 L1 范数在目标函数优化过程中对特征变量的收缩效应, 使之能根据样本数据自动识别与因变量最为相关的解释变量, 并将不具预测力的变量系数收缩为零, 从而构建出稳健且具有良好可解释性的模型 (Belloni et al., 2014)。考虑到农业生产极度复杂, 该特性对于农经研究中常见的高维数据具有重要意义。在实际分析中, 研究者往往面对大量潜在的解解释变量, 例如, 农产品价格预测模型中可能同时涉及气象指标、农资投入、政策干预、土地条件乃至宏观经济波动等多个维度。传统线性回归在变量数量接近或超过样本容量时往往无法有效估计, 而 LASSO 能在不完全舍弃线性建模逻辑的基础上, 以数据驱动的方式实现模型压缩, 从而在保持可解释性的同时提升稳定性与预测性能。

此外, 以 LASSO 为代表的正则化估计方法可结合交叉验证、信息准则等方法, 优化惩罚参数的选取, 进一步增强模型的泛化能力, 避免因过度压缩或欠拟合而损失有效信息。在处理农经领域实际问题时, 如识别影响农民支付意愿的关键因素等, LASSO 提供了一种灵活的变量选择与建模路径。其高维数据下对变量选择的自动化、对共线性问题的缓解及对过拟合风险的控制, 使其成为连接传统计量模型与现代机器学习技术的重要桥梁, 拓展了线性回归模型在农经研究中的应用边界 (王晓兵等, 2024)。

(三) 从规则划分到非线性结构识别：树模型与集成算法

在农经研究中, 自然条件、经济投入与社会网络等变量往往蕴含着极其复杂的非线性关系。在面对变量间关系复杂的农经研究问题时, 依赖先验假设的传统计量方法有一定的局限性。机器学习的树模型方法, 采用非参数的数据驱动策略, 凭借递归划分将样本空间分割为多个互不相交的局部区域, 从而对复杂关系进行刻画, 这一特性使树模型在非线性和非线性效应关系建模方面具有天然优势。然而, 尽管单棵决策树能够实现低估计偏差, 但其对训练数据中的噪声极为敏感, 容易陷入高方差的过拟合困境。为了化解这一难题, 以随机森林 (Random Forest, RF) 为代表的集成学习算法通过组合多棵决策树, 降低预测方差, 进一步提升模型的泛化能力。

随机森林算法的核心机制在于创新性地采用了“双重随机化”策略: 其一, 在样本层面, 通过自助法重采样, 为每棵树生成差异化的训练集; 其二, 在变量层面, 每次分裂节点时不再遍历所有变量, 而是随机选取一个子集用于寻找最优划分点。这种“双重随机化”机制削弱了单棵树之间的相关性。它在保留树模型捕捉高度非线性关系优势的同时, 通过集成均值化显著抑制了整体预测方差, 从而一定程度上缓解了传统模型在函数形式设定上的局限 (Rigatti, 2017)。因此, 对于给定的输入特征 x , 随机森林的最终

预测值由 B 棵树的预测结果加权平均或投票共同决定。具体公式如下：

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (4)$$

(4) 式中， B 为决策树数量， $T_b(x)$ 表示第 b 棵决策树对 x 的预测，其中每棵树 T_b 都是在一个重采样数据集上构建的，并且在每次分裂节点时，仅从特征全集中随机选取一个子集用于划分。与传统线性模型相比，树模型的最大优势在于不依赖先验设定的函数关系，而是以数据驱动方式自动学习输入变量与目标变量之间潜在的非线性关系 (Biau and Scornet, 2016)。每棵树通过一系列条件判断将样本空间分割为多个区域，从而对复杂的边界形态进行逼近。例如，在图1所示的谷物价格预测场景中，随机森林模型能够灵活整合积温、种植面积和替代品价格等信息。在无须预设函数形式的前提下，模型通过多棵决策树的集成与投票或平均机制，自动捕捉复杂的高阶交互特征，从而在复杂的自然与市场条件下实现对谷物价格的精准预测。

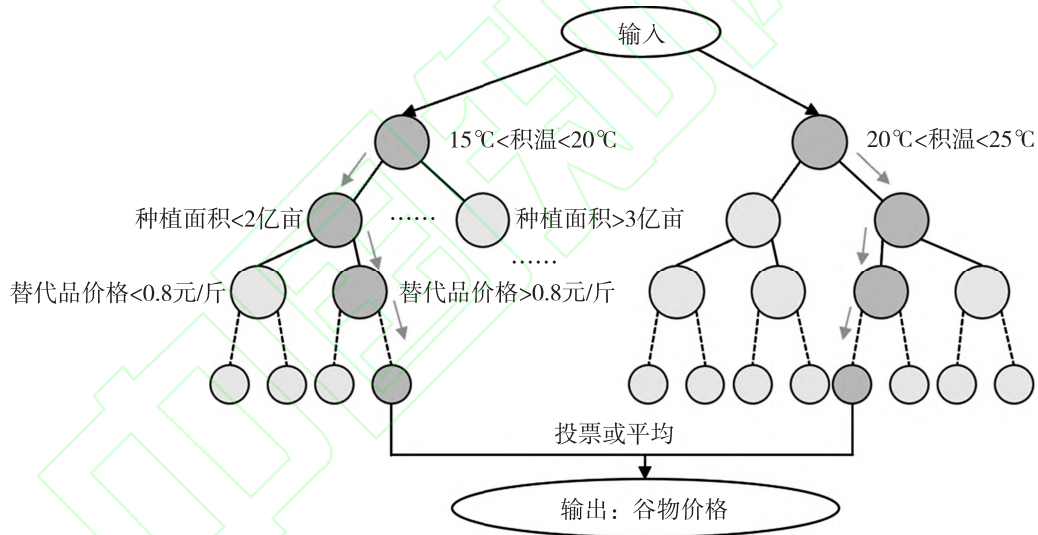


图1 随机森林模型预测谷物价格示意图

树模型通常被视为“黑箱”，其递归划分所构成的高度非线性结构难以揭示各解释变量对预测结果的作用机制。近年来，以 SHAP (shapley additive explanations, SHAP)、PDP (partial dependence plot, PDP) 等为代表的模型解释工具为解决这一问题提供了新路径。SHAP 通过遍历所有特征子集，将整体预测分解为各输入变量的平均边际贡献；PDP 则在保持其余特征分布不变的条件下，对目标特征进行插值并计算预测条件期望，进而描绘出该特征的响应曲线。凭借“可解释机器学习方法” (刘媛等, 2025)，研究者得以较为系统地刻画输入变量的边际效应及其高阶交互作用，为机器学习模型的参数赋予实证含义。在农经研究的多个关键领域，包括信用风险 (彭艳玲等, 2025)、城乡融合 (齐秀琳和汪心如, 2024) 等方面，树模型均展现出良好

的预测性能与适应性。它们作为连接传统统计方法与深度学习之间的重要中介，为农经研究提供了一种兼具灵活性与稳健性的建模路径，在数据复杂性不断提高的背景下，正逐步成为农经实证分析中的重要工具。

（四）复杂异构数据建模工具：深度学习模型

在农经研究中，诸如农产品价格波动、作物产量影响因素、农户行为决策等问题，往往伴随着高维的输入变量、复杂的非线性关系以及多层次的决策逻辑。传统线性模型与浅层算法由于依赖明确设定的函数结构，常难以捕捉这些变量之间深层次的相互作用与演进过程，尤其是在面对来自遥感、物联网、文本信息等非结构化数据源时表现出一定局限。为此，深度学习模型以其较强的特征学习能力和非线性函数逼近能力，为此类复杂问题的建模与预测提供了一种数据驱动的新范式（Zheng et al., 2023）。典型的深度学习算法包括多层前馈神经网络（multi-layer perceptron, MLP）、卷积神经网络（convolutional neural network, CNN）、循环神经网络（recurrent neural network, RNN）、长短期记忆网络（long short-term memory, LSTM）以及近年来表现突出的Transformer架构。这些模型通过引入多层神经元与非线性激活函数，模拟复杂的函数映射结构，在理论上具备逼近任意可测函数的能力。图2展示了如何利用深度学习算法处理高维非结构化的卫星遥感数据，将卫星图像转化为可用于农经研究的耕地分布与面积特征，从而对微观地块信息进行精准提取。

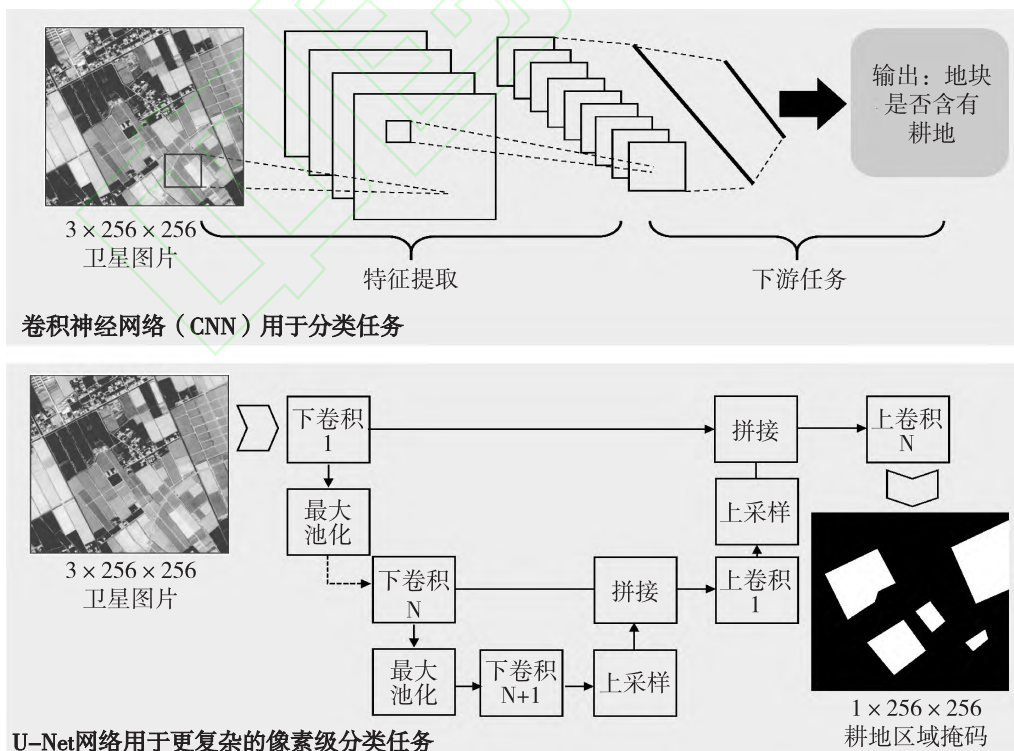


图2 卷积神经网络与U-Net网络在耕地识别中的应用

图神经网络（graph neural networks, GNNs）作为深度学习框架在非规则图结构上的延伸，以消息传递机制为核心，通过在多层网络结构中递归地聚合节点邻域信息，精准捕捉空间中的复杂关系。Fan et al. (2022) 提出将图结构与时序网络结合，以学习农作物产量与投入之间的复杂依赖关系，进而提升预测性能。这些方法拓展了深度学习在农经研究中的边界，为农户行为建模、作物产量预测、政策文本分析等任务提供了新的技术路径。近年来，以 Transformer 架构为基础的大语言模型逐渐在农经研究中显现出重要的应用潜力。Tzachor et al. (2023) 指出，大语言模型能够辅助农业推广服务，通过精炼科学知识、生成定制化建议和多语言交互，为农户提供针对性的农业技术指导与决策支持。这种基于大语言模型的农业推广服务模式，尤其适用于难以被传统推广体系覆盖的小农户。

深度学习模型具备适配多模态数据与多目标协同优化的优势，能够契合农经研究中多源输入与多源输出的复杂建模需求。第一，适配多模态数据，可同时处理图像、文本等多源输入。在遥感图像、政策文本与市场数据融合建模等场景中，深度神经网络能够实现跨数据类型的训练，构建统一的信息表征空间，从而提升模型的预测精度与泛化能力 (Paolo et al., 2024)。第二，支持多任务学习，具备联合输出能力。面对农业系统中相互制约的多重目标，传统计量模型通常只能针对单一因变量进行估计。而多任务深度学习模型通过底层的参数共享机制，能够在同一个架构内联合训练，并且同时输出多个相关联的结果。这种机制不仅利用了任务间潜在的关系来提升整体估计效率，更使得农经研究能够突破单一变量的局限，为实现复合系统多目标权衡提供方法支撑。

然而，深度学习模型在农经研究中的应用仍面临多重瓶颈。一方面，深度学习模型的训练依赖海量训练数据，农业领域有限的样本规模容易导致模型结果过拟合，进而限制其泛化能力；另一方面，深度学习模型具有典型“黑箱”特征，导致其难以像传统模型那样清晰地呈现解释变量的边际效应与经济学含义。因此，本文认为，机器学习的优势主要体现在预测、复杂模式识别与多目标评估方面，在强调因果识别的研究中，其作用更多是传统方法的补充而非替代。尽管目前已有研究运用可解释机器学习技术，对各解释变量在模型预测中的相对贡献进行量化，但这些解释方法本质上仍建立在变量间的统计相关性上，无法替代农经研究中严谨的因果推断。

四、重点研究领域机器学习的实践成效与前沿探索

农经学科的研究领域横跨自然、经济、社会三大系统，自然系统的气候与资源禀赋、经济系统中的农业生产与技术变迁、社会系统中的乡村治理与农民幸福感等共同塑造着“三农”发展的格局。在中国农业农村现代化转型的背景下，三大系统内部以

及系统之间的联系日益紧密，中国“三农”实践的丰富数据为构建具有本土特色的农经理论提供了肥沃土壤。本文以气候变化与资源环境、农业生产与技术进步、农村劳动力转移与农业转移人口市民化三个重点研究领域为例，探讨如何融合机器学习技术，展示机器学习推动农经研究的典型价值与方法路径。

（一）研究领域一：气候变化与资源环境

农业是气候变化最直接的作用对象。气候变化通过影响农业，进一步对社会经济产生深远影响（Auffhammer and Schlenker, 2014）。传统农经研究在探讨气候变化问题时，大多依赖官方提供的气象数据，但这些数据通常来自空间覆盖不完全的监测站点，因此，传统计量模型在分析气象变量之间的复杂非线性关系时存在一定局限。近年来，机器学习方法在气候变化与资源环境研究领域展现出多层次的赋能潜力。Chen et al. (2024a) 探讨地表臭氧对中国农业全要素生产率的影响，但由于全国地表臭氧监测的历史数据稀缺且空间覆盖不足，传统研究难以构建长时段的精细污染暴露指标。作者采用LightGBM、XGBoost与Super Learner等机器学习模型，训练县级地表臭氧模型，并利用其预测能力推算数据缺失年份的时空连续臭氧浓度。机器学习在该研究中显著提升了关键暴露变量的时空精度，使原本无法展开的因果识别成为可能。Eyring et al. (2024) 的研究则系统探讨了如何利用机器学习重塑气候模拟工具。该研究将机器学习嵌入气候变化研究已有框架之中，改进了极端事件的量化指标。目前已有学者应用随机森林、人工神经网络、深度时序模型等机器学习方法开展区域干旱预测、极端天气影响评估和全球气候模拟等，利用更加精细的微观数据和更加灵活的建模方式揭示气候变化对资源环境与农业经济的影响机制（Dhillon et al., 2024）。

未来研究可从以下两个方向进一步推进：第一，突破原有的数据框架，推动多模态信息融合。利用机器学习技术融合多模态数据，结合卫星遥感、地面物联网传感器、市场价格与政策文本等多源信息，并利用图神经网络、迁移学习与自监督学习，从数据层面挖掘新的气候与资源环境问题，拓展农经的研究边界。第二，深化机器学习与计量模型的融合。未来可在计量分析框架下引入LASSO与因果森林，对高维气象协变量进行特征筛选与异质性处理效应估计。同时将深度神经网络嵌入动态规划等结构计量模型，灵活逼近农户适应决策函数，为极端气候风险管理、农业保险和绿色转型政策制定提供更可靠的量化依据。

（二）研究领域二：农业生产与技术进步

农业生产体系同样面临高维动态和高度异质性的复杂情景，传统计量方法在多源非线性特征挖掘和预测问题中略显不足。随着数据多样化与算力提升，机器学习正逐步展现其优势：在投入、产出层面，相关研究结合双重差分法与机器学习，衡量外生冲击对农业投入强度和产量的影响（Grell et al., 2021；Deininger et al., 2023）；在农

产品市场预测方面，相比传统的ARIMA模型，卷积神经网络、循环神经网络以及长短期记忆网络表现更优（于晓华和刘爽，2024）。例如，农业生产效益研究中至关重要的农产品价格问题，由于多重因素影响，呈现出极强的非线性和高波动性。Paolo et al. (2024) 使用机器学习方法，将农产品价格、能源成本、汇率、天气指数等多维变量纳入预测模型，利用LSTM处理长时间序列的优势对农产品价格走势进行预测，发现其预测结果优于计量预测模型。

机器学习技术未来在该领域的应用可以关注以下三个方向：首先，与传统问卷调查数据或统计数据相比，机器学习能够整合遥感影像、物联网传感器、金融交易记录及高频气象资料，构建覆盖投入和产出全流程的时空大数据。未来研究可以结合深度学习和自监督学习技术，融合高维异构信息，对耕地管理强度、化肥农药施用量、碳排放等指标进行更客观、精细的测算，深化农业生产与农业技术进步的相关研究。其次，使用机器学习方法进行粮食生产的精准预测。可在全国尺度上，通过融合卫星影像、作物生育期图像与政策情境模拟等输入，以Transformer-TCN混合模型预测主粮产量和种植结构跨季、跨域的动态演化，为国家粮食安全提供决策建议。最后，机器学习也可用于辅助生产率的测算，主要思路在于突破传统索洛余值法、数据包络分析法和随机前沿分析的同时生产函数假设，利用随机森林、深度神经网络与半参数随机前沿等模型对不同规模和经营水平的主体进行分组，再在各子样本内部分别估算全要素生产率，从而放宽统一生产函数的限制；随后借助贝叶斯模型平均或因果森林，对多种测算结果自适应加权汇总，以减轻模型设定偏差带来的误差。同时结合SHAP等可解释机器学习技术，识别边际生产效率的关键驱动因素，为差异化技术推广和精准补贴政策提供更具针对性的量化依据。

（三）研究领域三：农村劳动力转移与农业转移人口市民化

农村劳动力转移与农业转移人口市民化是农经的重要研究主题，但其内在机制复杂，涉及多重因素。近年来有学者尝试引入机器学习方法以更全面地识别其背后的关键影响因素。刘媛等（2025）聚焦农民工进城落户意愿低这一现实问题，但由于落户意愿受多种因素共同驱动，传统计量模型依赖预设线性结构，难以在高维环境下充分识别变量的重要性，为了克服这一问题，该研究利用XGBoost构建模型，并借助SHAP对变量贡献度和作用方向进行解释，从而在保持预测性能的同时揭示关键机制，展示了可解释机器学习方法在识别影响落户意愿关键因素方面的优势。齐秀琳和汪心如（2024）则以农业转移人口市民化水平为研究对象，进一步展示了机器学习在处理复杂系统指标方面的优势。该文首先构建综合市民化指数，然后在解释变量中引入多个层面的特征，使用随机森林、GBRT、XGBoost等预测性建模来考察农业转移人口市民化水平的影响因素。相较解释性建模，机器学习预测性建模不预设模型的具体形式，能够更准确地捕捉市民化水平的最主要影响因素，体现了机器学习在处理多维指

数构建、高维特征筛选和异质性揭示方面的优势。

总体来看，机器学习方法为理解复杂迁移行为提供了新的分析工具，特别是在变量重要性识别、非线性结构刻画和多维指标构建方面展现出一定优势。但是，就农经研究的需求而言，其应用仍存在较大拓展空间。未来研究可从以下方向深化：第一，推动机器学习与结构模型的结合，将可解释性机器学习用于前期变量筛选、异质性识别和高维状态变量构造，再将核心特征嵌入结构模型中，以提高政策效应估计的稳健性。第二，借助因果机器学习方法系统识别不同群体在政策情境下的异质性处理效应，为因人施策、因城施策的市民化政策提供更加精准的证据。通过上述方向的推进，机器学习有望在未来被更系统地纳入农经学科的理论建构与政策评估体系，为理解和推动农业转移人口市民化提供更坚实的分析基础。

五、典型研究场景中机器学习的比较优势与嵌入逻辑

农业生产的根本特征在于经济再生产与自然再生产相互交织，其相关研究领域横跨自然系统、经济系统、社会系统三大系统。当前农经的研究边界不断延展，2024年国务院学位委员会将农林经济管理一级学科下的二级学科增加到6个，包括农业经济与管理、林业经济与管理、食物经济与管理、自然资源管理、农村发展、农商管理。然而，面对如此复杂的复合系统，立足于经济系统的传统计量方法在解释跨学科深层机制时存在一定的局限。如图3所示，哪些研究场景适合引入机器学习方法，以利用其在政策效应识别与动态预测、跨模态信息融合与建模、多任务学习与综合评估方面的优势，成为亟待探索的命题。本节将立足农经学科特质，系统剖析机器学习在典型场景下的嵌入逻辑与比较优势。

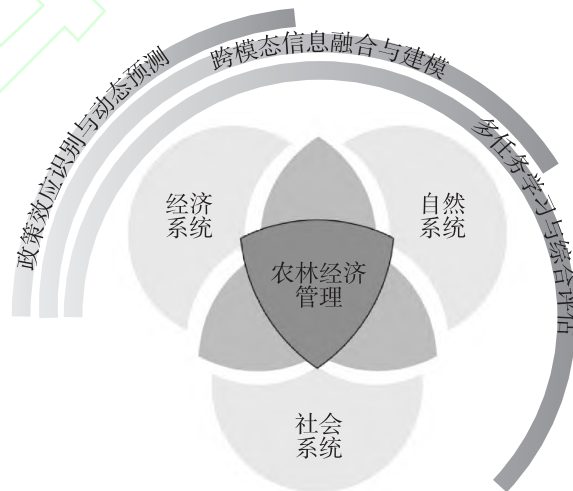


图3 农经学科特征及机器学习融合路径

（一）研究场景一：政策效应识别与动态预测

目前农经研究立足于经济系统，已形成相对完备的方法体系，传统方法被广泛认可和使用。然而，农业生产受自然条件、市场波动和政策变化等多重因素影响，观测单元之间存在较大的潜在异质性，且往往涉及复杂的高维变量交互。面对此类由多重复杂因素交织而成的研究问题，传统计量方法虽能提供基准分析，但在处理海量异构数据和捕捉非线性规律时往往力不从心。相比之下，机器学习凭借其在数据挖掘和函数形式上的高度灵活性，为突破传统方法的局限开辟了新路径。具体而言，机器学习主要在政策效应识别和动态预测两个方面，对现有的农经研究形成了有益补充。

在辅助政策效应识别上，一方面，机器学习提供了数据驱动的关键变量选择方案（Bollen, 2012），通过正则化机制有效克服了因果推断中难以选择工具变量和遗漏控制变量的挑战。例如，王晓兵等（2024）在研究农村食物市场对农村居民营养健康的影响时利用IV-LASSO方法、综合工具变量回归和LASSO方法的优势，实现对结果变量的高质量预测。同理，在处理样本选择偏差时，机器学习能够通过双重选择算法从复杂变量中精准识别关键混杂变量。Wagner and Zenker（2021）基于泰国小农户的面板数据，运用Imbens（2015）提出的半自动机器学习方法控制稻米价格保险政策中的选择偏误。该半自动机器学习方法有效减少了自选择问题，揭示了政策对产量、投资行为与收入结构的多重影响。

另一方面，在农经研究中，异质性处理效应日益受到关注，特别是在评估农业政策或补贴制度对不同类型农户产生的差异性影响时，其识别尤为关键。传统方法往往使用交互项验证异质性因果效应，但这依赖于研究者的先验假设，难以捕捉复杂的非线性异质性。相比之下，以因果森林为代表的方法，能够在保持模型结构灵活性的同时，基于数据驱动的方式划分样本，估计个体层面的条件平均处理效应（郭峰和陶旭辉，2023；Chen et al., 2024b）。这种方法特别适合揭示农业政策异质性影响，从而避免平均效应掩盖真实的政策价值。例如，Carter et al.（2019）在评估尼加拉瓜农村商业发展项目时，利用广义随机森林方法研究发现，尽管项目的总体平均效应较小，但基线条件较差的家庭从中获得了显著收益，这一发现强调政策精准靶向的重要性，避免了因平均效应不显著而认为项目无效的问题。类似地，胡尊国等（2022）利用因果森林方法评估了区域倾斜性政策的异质性，发现旨在协调平衡发展的政策在中国南方地区的效果显著优于北方地区。

农经学科定位决定了其研究不仅要立足于事后的政策效应评估，更需要在事前做出即时响应与动态预测。以研究农产品价格为例，传统计量范式更关注其中的影响因素和内在机制，强调理论的先验性和模型的解释能力；而机器学习在研究农产品价格问题时，则呈现出不同的研究思路，更加强调预测精度和数据驱动，较少关注因果关

系的细节机制 (Mullally et al., 2021)。首先, 针对具有高维特征和非线性波动的预测任务, 深度学习模型具有良好的表现。例如, Paolo et al. (2024) 通过纳入谷物等产品的期货价格和能源价格等特征, 构建了复杂的长短期记忆—循环神经网络 (LSTM-RNN), 以预测农产品未来的价格走势。虽然这种方法难以对单个变量的因果效应给出明确解释, 但在预测中具有较高的准确性, 为农业市场主体做出及时有效的决策提供了有力支持。此外, 支持向量机、随机森林、梯度提升树等, 因其良好的预测性和可解释性, 在消费者行为分析等场景中得到广泛应用。综上所述, 机器学习在政策效应评估和动态预测方面有效弥补了传统研究方法的不足, 但在结果可解释性和经济机制刻画方面仍存在短板, 需要与计量经济学方法深度融合, 共同构建更加完整且可信的分析范式。

(二) 研究场景二：跨模态信息融合与建模

若仅将分析局限于经济系统内部, 农经易被简化为一般应用经济学的子领域, 难以充分体现其多元学科特质。农业生产的本质是经济再生产与自然再生产相交织, 这意味着农经研究需要打破自然学科与社会学科之间的壁垒, 使自然系统的高维信息能够被量化并纳入经济分析框架。例如, 在评估化肥减量政策对农业产量的影响时, 作物产量的形成并非简单的线性投入产出过程, 而是需要融合土壤禀赋、气象条件、要素投入与管理决策等因素。然而, 传统计量经济学方法高度依赖结构化统计数据, 在处理非结构化自然信息时面临着天然的方法壁垒。事实上, 在自然系统研究领域, 机器学习已建立起较为成熟的技术范式, 这些前沿应用不仅拓展了观测数据的维度, 更为农经研究提供了宝贵经验与方法论启示。

农学领域在机器学习应用上的突破, 为农经研究提供了重要启示。农学研究通过机器学习方法解析基因与环境的相互作用及作物生长对气候的动态响应, 将传统的生产过程转化为可量化可预测的物理约束。例如, Guilpart et al. (2022) 利用多种机器学习方法建立气候和大豆产量之间的数据驱动关系, 预测 2050 年和 2090 年的大豆产量, 帮助农民、政策制定者和农业企业重新组织生产区域的分布。Liu et al. (2024) 则采用机器学习方法结合全球数据集, 将农田研究结果扩展到全球稻田, 研究发现稻作区基于有机替代的增产减排综合水肥管理策略, 为协同实现全球水稻产能提升和净碳减排提供了基于自然解决方案的路径。农学前沿的应用启示农经学者, 机器学习能够更加准确地将积温、病害等自然要素转化为经济变量, 从而进一步深化对农业的刻画。进一步地, 机器学习充当了连接微观作物生长过程和宏观经济决策的桥梁, 使得农学和农经的融合成为可能。例如, 有研究者将机器学习拟合的高精度非线性生产函数作为约束条件, 嵌入农户经济决策模型中 (Chang et al., 2023)。

在遥感领域, 机器学习已成为地表覆盖变化监测、作物类型识别等任务的核心工

具。利用卷积神经网络、随机森林等模型，研究者能够从多时相、多光谱甚至雷达影像中自动提取特征，实现对作物分布、播种结构、耕地撂荒和极端气候冲击的动态识别。Liu et al. (2025) 利用随机森林模型揭示了干旱灾害对次年春季物候的滞后影响机制，为评估气候变化的跨期生态后果提供了科学依据；Brandt et al. (2024) 则采用深度学习方法在次大陆尺度上自动识别了约6亿株农田树木并进行长期追踪。这些前沿应用通过将传统的非结构化影像数据转化为可量化的经济代理变量，极大地拓宽了农经的研究边界。结合机器学习在遥感中的应用，能够为在缺乏微观调查数据的背景下开展资源管理评估开辟新路径。

更关键的是，机器学习赋予了农经研究突破传统数据限制的能力，使研究者能够追踪长期演变趋势，并在更精细的时间窗口内捕捉高频动态变化，从而为揭示复杂的机制提供全时空视角的实证支撑。例如，Lobell et al. (2020) 关注非洲小农作物产量和生产函数估计，但传统研究中经常存在农户回忆偏误或瞒报产量，且非洲调查数据获取成本极高。为解决这些难题，作者使用机器学习方法结合遥感影像，训练遥感产量模型。结果表明，基于机器学习估算的遥感产量在与地块实测收获数据对比时表现出较高精度。这一研究也表明，相较仅依赖传统问卷数据的做法，遥感结合机器学习生成的产量指标能够减轻测量误差带来的偏差，使对投入边际产出、土地生产率和技术效率的估计更加精准，从而为小农生产行为分析和农业政策评估提供更可靠的实证基础。

农学、遥感、环境科学等学科，为农经的跨领域研究提供了多模态的观察视角。然而，如何在统一框架下综合利用图像、传感器数据与统计数据等不同口径的信息，成为一个关键的难题。进一步地，机器学习跨模态表示学习和自监督预训练方法使研究者可以在缺乏精细人工标签的情况下，从多种信息源中自动提取结构化表征。其基本思路是利用配对关系或掩码建模任务，让模型在训练中同时学习各模态的语义表示，并将其映射到统一向量空间，从而实现跨模态检索与联合分析 (Baltrušaitis et al., 2018)。以图像、文本联合模型为例，对比语言—图像预训练 (contrastive language-image pre-training, LIP) 方法可将同一遥感图像与其文字描述或地块属性视为配对样本，在训练过程中完成语义对齐，使模型能识别空间细节，也能涵盖政策背景与经营特征。这类技术通常依赖弱监督机器学习方法，即仅需要少量标注即可提供基本语义指引，相较完全无监督方法具有更好的可解释性，更适合嵌入农经研究情境。

(三) 研究场景三：多任务学习与综合评估

农经学科的边界早已超越单纯的农业生产，广泛延伸到农村发展、农商管理等多元领域。在自然、社会、经济复合系统中，评价体系存在多元化需求，不同研究目标之间往往难以完全趋同。例如，作物产量的最大化往往伴随环境污染增加，即使产量

和污染达到平衡，也并不意味着社会效益达到最佳。因此，传统的利润最大化、产量最大化等单一视角研究已难以满足对整体效益评估的需求，农经研究的复杂性需要多任务的研究视角，而机器学习能够为处理此类多目标优化问题提供方法论支撑。

传统的计量模型通常针对单一因变量进行独立估计，忽略了不同经济或生态指标背后潜在的共性机制。相比之下，多任务深度学习模型能够通过参数共享机制，在隐藏层共同学习不同任务间的潜在关联。在农业生产场景中，这意味着可以将作物单产、化肥使用量和农户收入作为相关任务进行联合训练。由于这些任务共享相同的土壤、气候及农户特征等约束条件，联合训练不仅能降低对样本量的需求，还能利用任务间的相关性作为归纳偏置，从而提高对边缘任务的预测泛化能力，实现生态后果、经济效益与社会效益的权衡分析。例如，Sun et al. (2022) 基于田间近地遥感时序数据，构建多任务深度学习模型，在同一模型中同时预测小麦产量与籽粒蛋白含量两个指标。结果显示，多任务模型在保证预测精度不逊于单任务模型的前提下，通过共享特征提取层提高了对不同性状的整体拟合，并能通过注意力权重解析不同时期植株结构与光谱特征对产量与品质的贡献。对农经研究者而言，这类多任务模型可以嵌入农户行为模型中，对不同政策情景下的污染排放、投入产出与农民幸福感进行模拟。

此外，机器学习与基于主体模型 (agent-based modeling, ABM) 的融合，为评估自然、经济、社会的整体效益开辟了新路径。随着农经研究边界向乡村治理与乡村社会网络延伸，这要求模型不仅能模拟微观农户的生产行为，还能体现出宏观的社会结构变迁。机器学习，特别是代理模型可以有效解决复杂 ABM 校准难的问题。通过训练深度神经网络来模拟农户在政策干预时的动态反应，研究者可以构建一个数字孪生的乡村社会系统。这样能够实现在虚拟实验室中预演政策对农民幸福指数、乡村贫富差距以及社区凝聚力等难以直接观测指标的长期影响，从而实现从单一经济指标预测向自然—经济—社会系统综合评估的范式跨越。

六、研究结论与未来展望

机器学习方法在数据处理、复杂模型构建、因果推断、预测与分类中的应用得到了农经学者的广泛关注。在“三农”转型持续推进的背景下，机器学习方法为农经研究提供了新的视角与方法支撑。必须指出的是，当前中国农经研究在应用机器学习方面仍处于起步阶段，本文从三个关键方向凝练了农经研究在新时代的转型趋势，以期对未来研究提供一定的理论启示与方法借鉴。

第一，从“政策评估为主、预测为辅”转向“前瞻性预测与政策评估并重”的趋势。传统农经研究主要依赖回溯性分析，通过计量模型识别外生冲击所带来的平均处

理效应，并辅以 ARIMA 等预测工具。然而，在农业风险高发、外部不确定性增强的现实条件下，以事后评估为主的模式难以满足快速响应和精准干预的治理需求。机器学习方法的融入，使事前预判、动态反馈的闭环系统逐渐成为可能，预测功能逐渐从辅助性角色转变为与政策评估并重的角色，从而进一步提升农经研究的现实意义与治理价值。

第二，从“基础统计”转向“跨模态多源信息融合”的趋势。如果仅将分析局限于经济系统内部，农经难以体现其经济再生产、自然再生产和社会再生产交织的本质特征。未来的研究将更加注重利用机器学习作为跨学科的连接器，将自然系统、经济系统和社会系统深度连接。这种跨模态的信息融合不仅能缓解农经研究的数据局限，更能从多维视角还原“三农”全景，从而更精准地把握“三农”问题的深层逻辑。

第三，从“单一任务”转向“多任务协同的系统评估”的趋势。传统计量模型往往聚焦单一目标变量，容易忽视农业系统的目标权衡。面对粮食安全、生态保护与农民增收等多重目标的协调难题，未来的研究将借助机器学习中的多任务学习框架，从局部均衡分析走向系统综合评估。通过构建能够同时处理多个关联任务的模型，研究者可以同步量化政策干预在环境污染、经济产出和社会公平等多个维度的效用，从而为制定兼顾效率与公平、发展与安全的复合型农业政策提供更为科学的决策依据。

总体而言，机器学习在推动农经研究范式转型过程中展现出巨大潜力，主要体现在以下三个方面：一是在研究逻辑上，推动先验假设与数据驱动深度融合，为精准刻画复杂非线性关系与挖掘潜在规律提供了新路径；二是在研究对象上，实现向非结构化信息的拓展，将难以量化的复杂农业变量转化为高精度指标，拓宽农经学科的研究边界；三是在决策维度上，融合多任务协同的综合评估与事前预测的优势，为跨学科、跨领域的复杂前沿问题构建了深度研究框架。但与此同时，也应清醒地认识到，机器学习并非万能工具，其本身仍存在诸多局限，在模型可解释性、因果识别、数据质量依赖性等方面仍有待完善。此外，机器学习无法独立完成理论建构、机制解释与价值判断，更不能替代严谨的因果逻辑推演。与之相比，传统研究范式虽在处理高维复杂数据方面不具优势，但在理论构建、机制分析和研究设计的逻辑严谨性方面仍具有不可替代的价值。因此，未来的农经研究应秉持“技术赋能、理论统领”的整合思路，既要发挥机器学习在预测、分类与扩展数据维度方面的能力，也要坚守传统范式在因果推断与理论解释中的优势，通过机器学习驱动，实现更具科学性、实践性与前瞻性的研究范式重构，为农业高质量发展和乡村振兴战略提供坚实支撑。

参考文献

方帅、刘知、程雨涵、李子璇，2025：《机器学习方法与企业管理研究的融合：回顾与展望》，《经济管理》第3期，第183-208页。

- 郭峰、陶旭辉，2023：《机器学习与社会科学中的因果关系：一个文献综述》，《经济学（季刊）》第1期，第1-17页。
- 洪永森、汪寿阳，2021：《大数据如何改变经济学研究范式？》，《管理世界》第10期，第40-55页。
- 胡尊国、顾金鑫、陈颖，2022：《“倾斜性”政策、生产部门变迁与南北地区发展差异——来自机器学习的因果推断》，《财经研究》第1期，第93-107页。
- 刘景江、郑畅然、洪永森，2023：《机器学习如何赋能管理学研究？——国内外前沿综述和未来展望》，《管理世界》第9期，第191-216页。
- 刘媛、熊柴、蔡继明，2025：《农民工进城落户的意愿为什么不高？——基于可解释机器学习方法的分析》，《中国农村经济》第4期，第20-41页。
- 罗必良，2020：《构建“三农”研究的经济学话语体系》，《中国农村经济》第7期，第2-23页。
- 彭艳玲、彭一杰、周红利、汪寿阳、蒋远胜，2025：《基于机器学习的农户农地经营权抵押贷款信用风险识别及其损失度量》，《系统工程理论与实践》第2期，第448-462页。
- 齐秀琳、汪心如，2024：《基于机器学习方法的农业转移人口市民化水平影响因素研究》，《中国农村经济》第5期，第128-150页。
- 钱文荣、郑淋议，2025：《中国农业经济管理学》，杭州：浙江大学出版社，第15-23页。
- 仇焕广、陈晓光、吕新业，2018：《农业经济研究的前沿问题与方法探讨》，《农业技术经济》第1期，第17-23页。
- 王芳、王宣艺、陈硕，2020：《经济学研究中的机器学习：回顾与展望》，《数量经济技术经济研究》第4期，第146-164页。
- 王晓兵、赵龙强、闵师、王悦，2024：《农村食物市场、食物消费与营养健康：基于IV-LASSO方法的检验》，《农业经济问题》第8期，第25-41页。
- 于晓华、刘爽，2024：《“没有免费午餐定理”与农业政策研究中的算法选择：以机器学习预测生猪价格变动为例》，《农业经济问题》第5期，第20-32页。
- Athey, S., 2017, “Beyond Prediction: Using Big Data for Policy Problems”, *Science*, 355(6324): 483-485.
- Athey, S., and G. Imbens, 2019, “Machine Learning Methods That Economists Should Know About”, *Annual Review of Economics*, 11(1): 685-725.
- Auffhammer, M., and W. Schlenker, 2014, “Empirical Studies on Agricultural Impacts and Adaptation”, *Energy Economics*, 46: 555-561.
- Baltrušaitis, T., C. Ahuja, and L. P. Morency, 2018, “Multimodal Machine Learning: A Survey and Taxonomy”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2): 423-443.
- Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen, 2012, “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain”, *Econometrica*, 80(6): 2369-2429.
- Belloni, A., V. Chernozhukov, and C. Hansen, 2014, “High-Dimensional Methods and Inference on Structural and Treatment Effects”, *Journal of Economic Perspectives*, 28(2): 29-50.
- Biau, G., and E. Scornet, 2016, “A Random Forest Guided Tour”, *Test*, 25(2): 197-227.
- Bollen, K. A., 2012, “Instrumental Variables in Sociology and the Social Sciences”, *Annual Review of Sociology*, 38(1): 37-72.
- Brandt, M., D. Gominski, F. Reiner, A. Kariryaa, V. B. Guthula, P. Ciaia, X. Tong, W. Zhang, D.

Govindarajulu, D. Ortiz-Gonzalo, and R. Fensholt, 2024, “Severe Decline in Large Farmland Trees in India over the Past Decade”, *Nature Sustainability*, 7(7): 860–868.

Carter, M. R., E. Tjernström, and P. Toledo, 2019, “Heterogeneous Impact Dynamics of a Rural Business Development Program in Nicaragua”, *Journal of Development Economics*, 138: 77–98.

Chang, Y., J. Latham, M. Licht, and L. Wang, 2023, “A Data-Driven Crop Model for Maize Yield Prediction”, *Communications Biology*, 6(1): 439.

Chen, X., J. Gao, L. Chen, M. Khanna, B. Gong, and M. Auffhammer, 2024a, “The Spatiotemporal Pattern of Surface Ozone and Its Impact on Agricultural Productivity in China”, *PNAS Nexus*, 3(1): 435.

Chen, Z., S. Dai, B. Gong, 2024b, “Learning-Based Analysis of the Impact of Environmental Attention on Food Industry Productivity”, *International Food and Agribusiness Management Review*, 28(5): 881–902.

Deininger, K., D. A. Ali, N. Kussul, A. Shelestov, G. Lemoine, and H. Yailimova, 2023, “Quantifying War-Induced Crop Losses in Ukraine in near Real Time to Strengthen Local and Global Food Security”, *Food Policy*, 115: 102418.

Dhillon, R., G. Takoo, V. Sharma, and M. Nagle, 2024, “Utilizing Machine Learning Framework to Evaluate the Effect of Climate Change on Maize and Soybean Yield”, *Computers and Electronics in Agriculture*, 221: 108982.

Eyring, V., W. D. Collins, P. Gentine, E. A. Barnes, M. Barreiro, T. Beucler, M. Bocquet, C. S. Bretherton, H. M. Christensen, K. Dagon, and others, 2024, “Pushing the Frontiers in Climate Modelling and Analysis with Machine Learning”, *Nature Climate Change*, 14(9): 916–928.

Fan, J., J. Bai, Z. Li, A. Ortiz-Bobea, and C. P. Gomes, 2022, “A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction”, *Proceedings of the AAAI conference on artificial intelligence*, 36(11): 11873–11881.

Grell, M., G. Barandun, T. Asfour, M. Kasimatis, A. S. P. Collins, J. Wang, and F. Guder, 2021, “Point-of-Use Sensors and Machine Learning Enable Low-Cost Determination of Soil Nitrogen”, *Nature Food*, 2(12): 981–989.

Guilpart, N., T. Iizumi, and D. Makowski, 2022, “Data-Driven Projections Suggest Large Opportunities to Improve Europe’s Soybean Self-Sufficiency under Climate Change”, *Nature Food*, 3(4): 255–265.

Gulati, K., K. Saha, and T. J. Lybbert, 2025, “Women’s Work and Agricultural Productivity Gaps in India”, *American Journal of Agricultural Economics*, 107(5): 1261–1289.

Huang L., S. Hsiang, and M. Gonzalez-Navarro, 2021, “Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs”, NBER Working Paper No. 29105, <https://www.nber.org/papers/w29105>.

Imbens, G. W., 2015, “Matching Methods in Practice: Three Examples”, *Journal of Human Resources*, 50(2): 373–419.

Jia, X., A. Khandelwal, D. J. Mulla, P. G. Pardey, and V. Kumar, 2019, “Bringing Automated, Remote-Sensed, Machine Learning Methods to Monitoring Crop Landscapes at Scale”, *Agricultural Economics*, 50: 41–50.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, “Deep Learning”, *Nature*, 521(7553): 436–444.

Liu, B., C. Guo, J. Xu, Q. Zhao, D. Chadwick, X. Gao, F. Zhou, P. Lakshmanan, X. Wang, X. Guan, H. Zhao, L. Fang, S. Li, Z. Bai, L. Ma, X. Chen, Z. Cui, X. Shi, F. Zhang, X. Chen, and Z. Li, 2024, “Co-Benefits for Net Carbon Emissions and Rice Yields through Improved Management of Organic Nitrogen and Water”, *Nature*

Food, 5(3): 241–250.

Liu, Y., Y. Zhang, J. Peñuelas, S. A. Kannenberg, H. Gong, W. Yuan, C. Wu, S. Zhou, and S. Piao, 2025, “Drought Legacies Delay Spring Green-up in Northern Ecosystems”, *Nature Climate Change*, 15(4): 444–451.

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray, 2020, “Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis”, *American Journal of Agricultural Economics*, 102(1): 202–219.

Ludwig, J., and S. Mullainathan, 2024, “Machine Learning as a Tool for Hypothesis Generation”, *The Quarterly Journal of Economics*, 139(2): 751–827.

Mullainathan, S., and Z. Obermeyer, 2017, “Does Machine Learning Automate Moral Hazard and Error?”, *American Economic Review*, 107(5): 476–480.

Mullally, C., M. Rivas, and T. McArthur, 2021, “Using Machine Learning to Estimate the Heterogeneous Effects of Livestock Transfer”, *American Journal of Agricultural Economics*, 103(3): 1058–1081.

Northcutt, C., L. Jiang, and I. Chuang, 2021, “Confident Learning: Estimating Uncertainty in Dataset Labels”, *Journal of Artificial Intelligence Research*, 70: 1373–1411.

Paolo, L. B., V. Alessandro, G. Cornelis, and S. Paolo, 2024, “Machine Learning to Predict Grains Futures Prices.”, *Agricultural Economics*, 55(3): 479–497.

Rigatti, S. J., 2017, “Random Forest”, *Journal of Insurance Medicine*, 47(1): 31–39.

Storm, H., T. Heckeley, K. Baylis, and K. Mittenzwei, 2024, “Identifying Farmers’ Response to Changes in Marginal and Average Subsidies Using Deep Learning”, *American Journal of Agricultural Economics*, 106(4): 1544–1567.

Sun, Z., Q. Li, S. Jin, Y. Song, S. Xu, X. Wang, J. Cai, Q. Zhou, Y. Ge, R. Zhang, J. Zang, and D. Jiang, 2022, “Simultaneous Prediction of Wheat Yield and Grain Protein Content Using Multitask Deep Learning from Time-Series Proximal Sensing”, *Plant Phenomic*, 9757948.

Tanriverdi, I., and İ. Batmaz, 2025, “AI-Driven US Drought Prediction Using Machine Learning and Deep Learning”, *Climate Dynamics*, 63(6): 1–24.

Tao, L., Z. Tian, W. Yang, and C. Zhang, 2024, “Media Coverage of Food Safety and Expenditures for Food Away from Home: Empirical Evidence from China’s Family Panel Studies”, *China Agricultural Economic Review*, 16(4): 871–888.

Wagener, A., and J. Zenker, 2021, “Decoupled but Not Neutral: The Effects of Counter-Cyclical Cash Transfers on Investment and Incomes in Rural Thailand”, *American Journal of Agricultural Economics*, 103(5): 1637–1660.

Wooldridge, J. M., 2010, *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press, 49–52.

Wuepper, D., W. A. Oluoch, and H. Hadi, 2025, “Satellite Data in Agricultural and Environmental Economics: Theory and Practice”, *Agricultural Economics*, 56(3): 493–511.

Zheng, Y., Z. Xu, and A. Xiao, 2023, “Deep Learning in Economics: A Systematic and Critical Review”, *Artificial Intelligence Review*, 56(9): 9497–9539.

Machine Learning-Driven Paradigm Shift in Agricultural Economics Research

CHEN Zimeng¹ SHEN Yuhui¹ JIANG Zhuoren¹ GONG Binlei^{1,2}

(1.School of Public Affairs, Zhejiang University;

2.China Academy for Rural Development, Zhejiang University)

Summary: Amidst the rapid development of rural areas and the escalating complexity of research inquiries, the limitations of traditional paradigms in agricultural economics, particularly in addressing high-dimensional structures and complex model specification have become increasingly prominent. Machine learning (ML), leveraging its advantages in deep data mining, non-linear modeling, and predictive analysis, offers a powerful technical complement and methodological expansion for the field. The research focuses on the driving role of machine learning in the paradigmatic innovation of agricultural economics research, elucidating its application logic across data acquisition, model construction and typical research scenarios by comparing it with traditional econometric methods. The research reveals that machine learning through its capacity for deep unstructured feature extraction and flexible non-parametric modeling, can effectively capture the intricate patterns interwoven within the natural, economic and social dimensions of agricultural production, thereby providing a robust extension to traditional research frameworks.

The research makes a theoretical contribution by analyzing the breakthroughs of machine learning in multi-source data acquisition and processing compared to traditional methods, reconstructing the data architecture of agricultural economics through satellite remote sensing, text mining, and sensor data to facilitate a transition from structured data to multimodal observations. Furthermore, the study demonstrates the comparative advantages of machine learning in high-dimensional variable selection, non-linear structure identification, and cross-modal information fusion through regularized estimation, tree models, random forests, and deep learning architectures. In key research domains such as climate change, agricultural production, and labor migration, machine learning not only enhances predictive accuracy but also deepens scientific understanding of heterogeneous treatment effects and policy mechanisms through tools such as Causal Forests and interpretable machine learning.

The study concludes that agricultural economics under the drive of machine learning is exhibiting three major transformative trends: shifting from retrospective policy evaluation toward a dual emphasis on forward-looking prediction and evaluation; transitioning from basic statistics to cross-modal multi-source information fusion; and moving from single-task analysis toward multi-task collaborative systematic assessment. This paper proposes that future research in agricultural economics should adhere to the principle of “technology empowerment guided by theory”, deeply integrating the predictive strengths of machine learning with the causal explanations of traditional paradigms. This integration not only advances the understanding of the methodological evolution of agricultural economics but also provides significant reference for fostering paradigm innovation and accelerating the construction of an independent knowledge system for Chinese philosophy and social sciences.

Keywords: Machine Learning; Agricultural Economics; Research Paradigm

JEL Classification: Q10; C45; O33

(责任编辑:陈静怡)