

# 机器学习和农业政策研究范式的革新<sup>\*</sup>

于晓华 (德国哥廷根大学农业经济与农村发展系 哥廷根 37073)

唐 忠 (中国人民大学农业与农村发展学院 北京 100872)

包 特 (新加坡南洋理工大学经济学系 新加坡 639798)

**内容提要:**传统的农业政策分析偏重于分析变量之间关系和模型参数的估计,而“机器学习”更注重对未来结果预测的准确性,而后者恰恰是农业政策分析的目的。“机器学习”具有庞大的数据收集和储存能力、强大的学习分析能力以及更智能化的语言分析能力等,所以“机器学习”会对农业政策研究带来革命性的影响。农业经济学界要关注“机器学习”的发展,在科研和教学中导入“机器学习”,结合传统的农业经济学分析方法,让农业政策的制定更加精准、更加科学、更加强大,政策沟通更加有效。

**关键词:**机器学习; 农业政策研究; 农业经济学; 预测; 计量模型

## 一、前 言

2018年是中国农业政策和农村发展历史转变的关键一年,中央一号文件的主题是“乡村振兴”。在经历改革开放后40年高速经济发展,中国农业占GDP的比重从1978年的28.2%下降到2017年的7.9%。农业小部门的趋势会进一步强化。城市化率也达到58.5%。传统农业政策侧重关注粮食安全以及农民收入和贫困问题。这两者在现在的中国已经基本得到解决(于晓华等,2012; Yu, 2018; 于晓华, 2018)。

现在农业和农村发展遇到一系列新的问题:农业生产成本上升、农业竞争力下降、进口和库存暴增、农业生产的生态环境压力巨大、农村老龄化迅速推进、乡村人口不断衰落。在这样的背景下,中央提出“乡村振兴”的伟大战略,指导今后几十年中国农村经济的发展走向(于晓华, 2018)。

乡村振兴是一个系统工程。一号文件提出的总要求为产业兴旺、生态宜居、乡风文明、治理有效和生活富裕。中国农业和农村处于一个剧烈动态变化之中,如何制定更好且具体的“乡村振兴”政策和措施,向学术界提出更高的要求。现实中存在多种政策模型,不同的政策模型可能得出差距很远的各种政策建议,如何找到最合适的模型一直是政策分析中的难题。

伴随着计算机计算和存储能力的飞跃发展,机器学习(Machine Learning)开始走到科学研究舞台的前沿。机器学习是人工智能的一个分支,它研究如何学习利用数据或者以往的经验,改善具体算法的性能(杨善林等,2004; Lesmeister, 2015)。机器学习给科学研究和休闲娱乐等各个领域带来的影响可能都是革命性的。

机器学习给农业技术和农业政策研究提供了新的强大的工具,也带来很多新的挑战。在农业技术领

<sup>\*</sup> 项目来源:国家自然科学基金项目“我国政府粮食的规模优化与政策评价——基于理性预期假设的模型构建、水平则算与福利估计”(编号:71673289),食品安全消费者行为与风险交流策略研(编号:71633005),以及中国人民大学重大规划项目“工业化以来主要国家农业发展模式比较研究”(编号:15XNLG01)。唐忠为本文通讯作者

域,世界主要发达国家开始利用人工智能发展精准农业,以克服农业成本上升和老龄化的压力,提高农产品质量和产量,降低农业对环境的压力。由于中国农业的平均经营规模很小,所以精准农业的应用发展滞后,但是其在研究方面正在奋起直追(金继运,1998;刘爱民等,2000;李忠武等,2001;赵春江等,2003)。

机器学习对农业政策研究带来的影响也是巨大的。传统农业政策研究过程一般是先收集数据,分析主要变量之间的关系,然后给出政策建议,提交给政策制定者讨论通过。考虑到计算机庞大的数据收集和储存能力,现在“机器学习”能够自我分析学习这些数据,发现传统分析方法所忽略的一些规律,并能考虑和模拟未来社会和经济的动态变化,然后提出更科学的政策建议。

传统的农业政策分析,更偏重于分析变量之间的关系,而机器学习更注重对结果的预测,而后者恰恰是农业政策分析的主要目的。相比较传统的政策分析,“机器学习”在农业政策研究的某些方面可能更加精准、更加科学、也更加强大。

农业经济学是一门应用科学。农业经济学界要注重机器学习这方面的发展,在科研和教学中导入“机器学习”,结合传统的农业经济学分析方法,让农业政策的制定更加精准、更加科学、更加强大。

农业政策的制定和实施需要很好的沟通。一个好的政策建议需要和社会各界,尤其是政府部门进行沟通。由于各个政策利益主体所掌握的信息和利益诉求不同,“机器学习”可以通过自然语言的学习和分析,为各个利益主体量身定制政策沟通的方法,使得政策建议能够被社会各界广泛接受并付诸实践。

## 二、机器学习的发展

“学习”是获取和传递人类知识的重要手段。“学习”在经济学(尤其是实验经济学)研究中一直是一个重要的课题。知识和信息的获取与市场均衡的达成过程,都是一个学习的过程。Camerer(2003)在其经典教科书《行为博弈》中总结了各种学习。这包括演化动态强化学习、信念学习、老练(预期性)学习、经验加权吸引力学习、模仿学习、方向学习以及规则学习等。经济学中对于“学习”的研究和机器学习中的“学习”有很多相类似的地方。

从计算机诞生后的1950年代起,人工智能就开始成为一个重要的研究对象。作为人工智能的一个重要分支,机器学习在1980年代后在人工智能界引起了关注,逐渐发展成为一个庞大的具有广阔应用前景的学科(杨善林等,2004)。

机器学习的核心目标是机器从过去的经验中学习出一般的规律,并能够准确分析和预测新的和未曾经历过的任务。“机器学习”一般分为监督式学习、非监督式学习以及介于两者之间的半监督式学习。Athey(2018)通过经济学的实际例子比较了他们之间的区别。非监督式学习主要是发现观测对象的聚类,得出他们相似处,所以有时也被称为“降维分析”。非监督式学习主要用来分析视频、图片或者文字。对经济分析而言,非监督式学习可以帮助生成变量。而监督式学习主要是通过给出一系列特征或者变量(X)来预测一个结果(Y)。正如Athey(2018)所强调的那样,监督式学习不是聚焦于预测,而是机器在X和Y都可以观测的情况下得到训练,然后用其他的X来预测Y。这和计量经济学比较类似。但是计量经济学强调估计参数,而监督式学习更强调于预测的准确性(Mullainathan等,2017)。

经过30多年的发展,结合统计学和数学规划等领域,机器学习已经发展出很多学习方法。这包括决策树学习、深度学习、强化学习、人工神经网络学习、贝叶斯网络学习、基因学习、聚类学习等方法(杨善林等,2004)。不管哪种学习方法,机器学习利用大数据作为过去的经验,结合数值模拟,学习效率在很多方面超越人的认知能力。

机器学习已经在很多领域带来革命性影响。比如在通过深度学习建立的Alpha Go在2016—2017年连续战胜多位世界围棋顶尖高手,使其在围棋界独步天下。Athey(2018)认为机器学习也会给经济学带来深远影响,会给经济学带来新的研究问题和新的分析方法。她认为主要的影响有这四

个方面:(1)机器学习并不会给经济学分析的识别问题带来新的帮助,但是会让模型的选择更加富有弹性;(2)机器学习可以很容易的比较各种实证分析模型;(3)经济学可以把模型“外包”给机器学习,让数据更能吻合模型;(4)机器学习要修正模型,以给估计结果提供有效的置信区间。

必须指出的是非监督式的机器学习对视频、图片和语言文字具有很强的学习和分析能力,这可以帮助更加清晰的了解政策各个利益主体的需求,从而实施有效的政策沟通。

农业作为国民经济的重要组成部分,农业经济学作为一门应用科学,机器学习给农业政策分析带来的冲击也必然是巨大的,而很多影响现在还没有体现。

### 三、机器学习对农业政策研究的影响

农业政策制定需要科学性和预测准确性。政策科学性的基础是数据的有效收集、模型的科学建立以及模型计算的精准性和效率性。在计量经济模型和统计模型已经逐步在政策分析中普及的今天,政策分析的科学性似乎已经建立。机器学习能够给农业政策分析带来什么新的东西呢?

于晓华(2014)认为计量经济分析需要注重数据和模型两个方面的准确性以及计算的经济性。本文在这里就对机器学习对政策研究中的数据、模型以及计算的影响做一个具体的分析。

1. 数据。随着计算机数据收集和储存能力的不断增强,数据来源的不断扩展。大数据已经渐渐成为经济分析的基础,其对政策分析的作用是多方面的(Einav等,2014;张浩然等,2004;Lesmeister,2015)。于晓华(2014)指出数据来源包括政府机关数据、调研数据以及网络数据等。机器学习能够给政策提供更多数据,并能整合各种数据资源。

首先,机器学习能够提供更多数据。一个简单的运用就是网络爬虫,它能搜集并整理网络相关数据,减少数据收集的成本,并增加数据的来源。其次,数据来源是多方面,机器学习可以通过学习分析整合不同的数据来源。既减少分析时间和成本,也为政策研究提供更多的维度和信赖的结果。一个很好的应用例子就是周洁红等(2016)运用网络数据挖掘技术和文本内容分析方法,对2009—2014年与中国猪肉质量安全相关的10160则新闻报道进行归类分析,从中总结食品安全事件中社会共治的3类主体——政府、生产经营者和社会监管力量,各自发挥的作用及存在的问题,并提出相应的政策建议。这样的研究数据收集成本低廉,更容易发现一些一般分析所难观察到的经济规律,提出的政策更具有针对性。第三,非监督式的机器学习能够分析视频、图片以及语言文字等非标准化数据,通过聚类等方法,使得数据“降维”,为政策分析提供新的变量以及更标准化的数据。但是,还存在由于这方面的研究一定的技术障碍,比如语言识别的精确性不够、现阶段这方面的应用还非常少见。最后,机器学习还可以通过蒙特卡洛模拟等方法提供更多的分析或者模拟数据,使得数据内容更加丰富,使其政策预测更加准确。

总之,机器学习能够提供更多的数据,把非标准化的视频、图片以及语言文字等“降维”为标准化数据,并能够整合不同的数据来源。这为政策分析提供了更全面更丰富的数据,制定的政策更具有针对性。具体到农业政策分析,机器学习可以收集古今中外农业政策的案例、成功和失败案例以及现有各种相关的微观宏观调研和统计数据。比如于晓华等(2017)就指出欧盟共同农业政策的历史发展过程和中国类似,所以其成败的经验值得中国农业政策制定者学习。

2. 模型和计算。相对于实证计量经济分析更关注于参数的估计,机器学习更加关注预测的准确性(Mullainathan等,2017;Athey,2018;Lesmeister,2015)。机器学习出发点是数据,然后找出最好的模型与之吻合。计量分析中如何权衡模型的表达性和吻合性,一直是个难点。但是机器学习可以找到很多很好的方法来权衡(Athey,2018)。

相对于计量经济学中模型的选择是预先设定的,机器学习的出发点是数据,通常是比较多个模型,从而找出对数据吻合最佳的模型;很多时候,机器学习的分析模型可能是非参数的或者是动态的。

这样, 机器学习的模型选择更有弹性, 对政策结果的预测也更加准确。对于政策研究而言, 最重要的不是模型的精美, 而是结果的准确。如果机器学习提供的政策预测更加准确, 那机器学习在政策分析中淘汰计量模型的可能性就很大。

现实政策分析中存在多种模型, 模型的分析结果通常会差异很大, 这导致政策分析存在很大的争议。但是, 机器学习可以通过自身强大的计算能力, 在农业政策分析时在自我的学习中分析过去各种农业政策案例的成败经验, 也可以对照各种变量建立联系, 或者通过概率模拟分析政策的各种效果。这样的政策建议会考虑政策历史路径, 参考其他国家的经验, 结合现实和未来的各种可能选择, 使得政策决策考虑更加全面、更加科学, 减少政策失误的可能性。

政策分析中一个很重要的应用是时间序列变量之间的因果关系。赵国庆等(2009)分析建国以来通货膨胀和粮价之间的格兰杰因果关系, 发现两者之间存在长期均衡。因为粮价过高会伤害消费者利益, 而粮价过低会伤害农民的利益。但是, 对于他们之间的因果关系还是存在很多争议。格兰杰因果检验的模型选择存在一个难点: 最优滞后长度的选择, 不适当的滞后长度会导致分析结果存在谬误。Lesmeister(2015)指出运用机器学习可以选出最优的滞后量, 使得格兰杰因果检验更精确。

我国的乡村振兴战略作为一个长远的系统工程, “到2020年, 乡村振兴取得重要进展, 制度框架和政策体系基本形成; 到2035年, 乡村振兴取得决定性进展, 农业农村现代化基本实现; 到2050年, 乡村全面振兴, 农业强、农村美、农民富全面实现”, 乡村振兴战略具体政策制定时, 需要考虑的变量非常多、政策内容非常庞杂、政策目标甚至可能存在冲突, 短期目标和长期目标的协调, 各个利益主体的协调等因素(全世文等, 2016)。传统的政策分析研究方法可能不够系统性, 机器学习可以给乡村振兴战略的政策制定提供很好的帮助, 使得具体政策的制定更加科学和精确。此外, 机器学习可以很好打破宏观模型和微观主体行为之间的界限。现有的大数据可以直接记录和保存个体微观的数据, 在政策的模拟的时候, 可以直接模拟到每个个体效应, 然后直接加总微观主体的效果得到政策的总体效果, 使得其宏观政策的偏差能够更小。

#### 四、机器学习对农业政策沟通的影响

农业政策的制定过程包括了政策研究、政策分析以及政策沟通。一个好的政策建议, 能够上升为国家政策, 得以实施, 还需要很好的政策沟通。

政策制定会影响很多利益主体。具体到农业政策制定, 利益主体包括: 政府、农民、消费者、粮食贸易商、粮食加工商、生产要素供给商、环保机构等。很多利益主体之间的利益目标并不是一致的, 如何和他们进行有效政策沟通就显得尤其重要, 这关系到政策的顺利通过和实施。全世文等(2016)和赵国庆等(2008)就指出建国后我国粮食价格的政府干预一直存在对农民利益和城市消费者利益的权衡, 粮价太高会伤害消费者利益, 粮价太低又会伤害农民的利益。如何通过沟通让各个利益主体之间能够求同存异, 保证政策实施显得至关重要。机器学习在政策沟通中有很多优势。

首先, 非监督式机器学习可以通过聚类法等学习相关的视频、图片、对话、以及文字等非标准数据, 获取各个利益主体的利益诉求, 这样在制定政策的时候可以考虑他们的利益, 使得政策更容易通过并实施。

其次, 中国的政策制定更多的是由上而下的过程, 政府在政策制定过程中的主导作用非常明显。如何使政策建议更能够让政府接纳, 这需要了解政府的政策偏好, 甚至对政策建议的写作风格都有一定的要求。机器学习能够对这些非标准的数据进行分析, 针对各个政策制定者的个人偏好定制不同的政策报告, 使得政策建议者和政策决定者之间的沟通更加顺利, 政策建议更容易被采纳。

最后, 政策被政府接纳通过后, 其实施也需要和社会各界沟通。机器学习能够很容易分析并反映

出各个政策主体的利益需求,从而可以通过视频和图片等手段对各个利益主体进行特别沟通。

现实中,已经有学者开始运用机器学习中的文本挖掘技术,对各种政府政策文件进行分析。韩国明等(2017)运用文本挖掘技术分析了1982—2016年的一号文件,发现政府对农业经营主题的关注一直在变化,但是农户制始终贯穿其中,农民的合作问题始终是决策层关注焦点。但是,该分析只是关注了政府的一号文件,社会对政策反馈等还缺乏深入分析。

以我国的乡村振兴战略为例,政策制定和实施所需要沟通的内容和对象非常多。乡村振兴的总目标为“产业兴旺、生态宜居、乡风文明、治理有效和生活富裕”。其利益主体不仅是农民和政府,也包括涉农产业、环保工作、以及社会文化工作等利益主体。有些时候,各个利益主体之间的利益可能是不一致的(全世文等,2016),比如产业兴旺和生态宜居之间可能存在利益冲突。如何让政策和各个利益主体有效沟通,机器学习可以提供很好的帮助,找出他们之间的共同点,使得政策沟通求同存异。

中国处于一个高速发展的过程中,政策处于动态变化之中(蔡昉等,2008;于晓华等,2014;于晓华,2018;Yu,2018),如何就政策变化和各个利益主体有效沟通,机器学习是一个很好的辅助手段。

## 五、总 结

伴随着计算机储存和计算能力的发展,大数据和机器学习开始走上历史舞台,经过30多年的发展,机器学习开始对科学研究和人民生活产生巨大且革命性的影响。

农业经济学是一门应用科学,而农业政策分析是农业经济学的核心内容之一。机器学习对农业政策研究的冲击是巨大且可能是革命性的。

传统的政策分析更注重模型参数的估计,机器学习更注重政策预测的精确性。它能够提供更多的数据,从数据出发,通过智能学习和庞大的计算能力,找到更吻合的模型,从而产生最精确的预测结果。政策分析本身更追求预测的精确性,这为机器学习在农业政策分析中广泛应用提供了最好的土壤。

机器学习对非标准数据也有很强的学习能力,比如视频、图片、对话以及文字等。机器学习通过“降维”学习为政策分析提供更多的标准数据,也能更好的在政策研究中分析各个利益主体的利益诉求,并为政策沟通提供很好的辅助手段。机器学习在进行农业政策分析时在某些方面超越传统的分析手法,所以在农业经济学的研究和教学中导入机器学习,也是不可避免的趋势。在专业课程设置的时候,对农业经济学专业的学生,要强化计算机编程以及统计学相关的训练,使得学生至少掌握一门机器学习相关的主流计算机语言,例如R(Lesmeister,2015)或者Python。

随着经济的高速发展,中国农业发展和农业政策处于一个快速的动态调整过程。2018年中央一号文件提出了乡村振兴战略,为今后的农业发展指明具体发展方向。但是具体的农业政策制定还需要更细致和科学的研究,需要权衡各利益主体之间的关系,权衡政策的短期目标和长期目标之间的关系,考虑农业和国民经济整体发展之间的关系等。精准而有效的农业政策制定以及与各相关利益主体之间进行有效沟通是一个巨大的挑战,机器学习有可能为制定有效且精准的农业政策,实现乡村振兴贡献力量。

现有的农业经济学者已经意识到机器学习对农业政策分析的重要作用,并开始将其导入到政策分析中,比如网络数据收集和文本分析等(周洁红等,2016;韩国明等,2017)。总体而言,机器学习在农业政策分析中还基本处于空白阶段。随着机器学习技术的飞快发展,它必然在不久的将来主动进入农业经济分析领域。农业经济学者必须要主动迎接这个挑战。

## 参 考 文 献

1. Athey Susan. "The Impact of Machine Learning on Economics," NBER Chapters, in: The Economics of Artificial Intelligence: An Agenda National Bureau of Economic Research, Inc. 2018

2. Camerer C. Behavior Game Theory: Experiments in Strategic Interaction. Princeton University Press. NJ. USA. 2003
3. Einav Liran and Jonathan Levin. The Data Revolution and Economic Analysis , in Innovation Policy and the Economy , Volume 14 , Leimer and Stern. 2014
4. Lesmeister Cory. Mastering Machine Learning with R , Packt Publishing Ltd , Birmingham 2015 B3 2PB , UK
5. Mullainathan Sendhil , and Jann Spiess. "Machine Learning: An Applied Econometric Approach. " Journal of Economic Perspectives , 2017 , 31 ( 2 ) : 87~106
6. Yu , Xiaohua. 2018. Engel Curve , Farmer Welfare and Food Consumption in 40 Years of RuralChina. China Agricultural Economic Review. 2018 , Vol. 10( 1 ) : 1~15
7. 蔡昉等. 中国农村改革与变迁: 30 年历程和经验分析. 格致出版社 , 2008
8. 韩国明 , 郭鹏鹏. 我国农业经营主体变迁的政策意图及其演化路径. 中国农业大学学报: 社会科学版 , 2017( 2 ) : 93~101
9. 金继运. "精准农业"及其在我国的应用前景. 植物营养与肥料学报 , 1998( 1 )
10. 李忠武等. "精准农业"及其在我国的研究应用. 地理科学 , 2001( 6 ) : 564~569
11. 刘爱民 , 封志明 , 徐丽明. 现代精准农业及我国精准农业的发展方向. 中国农业大学学报 , 2000( 2 ) : 20~25
12. 全世文 , 于晓华. 中国农业政策体系及其国际竞争力. 改革 , 2016( 11 ) : 131~138
13. 杨善林 , 倪志伟. 机器学习与智能决策支持系统. 科学出版社 , 2004
14. 于晓华等. 如何保障中国的粮食安全. 农业技术经济 , 2012( 2 ) : 2~9
15. 于晓华等. 欧盟农业改革对中国的启示: 国际粮食价格长期波动和国内农业补贴政策的关系. 中国农村经济 , 2017( 2 ) : 84~96
16. 于晓华. 如何正确运用计量经济模型进行实证分析——实证分析中的数据、模型与参数. 农业技术经济 , 2014( 7 ) : 4~16
17. 于晓华. 以市场促进农业发展: 改革开放 40 年的经验和教训. 长江产业研究院报告 , 2018
18. 张浩然等. 农业大数据综述. 计算机科学 , 2004( 11 ) : 387~392
19. 赵春江等. 精准农业技术体系的研究进展与展望. 农业工程学报 , 2003( 4 ) : 7~12
20. 赵国庆 , 于晓华 , 曾寅初. 通货膨胀预期与 Granger 因果性研究. 数量经济与技术经济研究 , 2009( 4 ) : 29~39
21. 周洁红等. 社会共治视角下猪肉质量安全治理问题研究——基于 10160 个猪肉质量安全新闻的实证分析. 农业经济问题 , 2016( 12 ) : 6~15

## Machine Learning and Renovation of Agricultural Policy Research

YU Xiaohua , TANG Zhong , BAO Te

**Abstract:** Traditional agricultural policy analysis pays more attention to the relationship between variables and estimation of model parameters , while "machine learning" focuses on the accuracy of predictions , which is precisely the purpose of agricultural policy research. Because "machine learning" has enormous data collection and storage capabilities , strong learning and analysis capabilities , and more intelligent language analysis capabilities , it could have a revolutionary impact on agricultural policy research. The agricultural economists should pay attention to the development of "machine learning" , introduce "machine learning" in scientific research and teaching , and combine "machine learning" with traditional agricultural economic analysis methods to make agricultural policy formulation more precise , scientific and powerful , and make policy communication more efficient.

**Keywords:** Machine learning; Agricultural policy research; Agricultural economics; Forecasting; Econometric model

责任编辑: 吕新业