

机器学习对经济学研究的影响研究进展^{*}

黄乃静 于明哲

内容提要:机器学习与经济学研究的融合将改变传统经济学的研究方式。本文就机器学习对经济学研究的影响进行了较为系统的梳理,着重分析了机器学习在大数据背景下对丰富经济数据多样性的贡献,机器学习对经济预测准确性的改进作用,以及机器学习在估计平均处理效应、处理效应异质性和结构模型等因果推断中的应用,对这些领域的重要研究进行了比较详细的介绍。在阐述机器学习优势的同时,本文也指出在经济学研究中使用机器学习方法可能存在的局限性,并对未来的研究方向进行了展望。

关键词:机器学习 大数据 预测 因果推断

随着技术的进步,机器越来越在工作和生产活动中扮演起了人的角色。例如,Facebook 能在照片中识别出人脸,Siri 能够理解人的语言,谷歌能够直接翻译网页上的文字等。上述新兴科技实际上都依赖于机器学习(Machine Learning)技术的发展。什么是机器学习?从狭义的角度理解,所谓的机器学习就是一类算法,属于计算机科学的一个子领域(Alpaydin, 2014),但从广义的角度来看,它又是一种综合了计算机科学、工程学、统计学等多学科知识的新技术。机器学习的方法不仅被广泛地应用于理工医学等领域,并且越来越受到社会科学研究者的重视。许多经济学家都认为机器学习会对经济学的发展产生广泛且深远的影响,以至于在不久的将来,机器学习将改变经济学的研究方式(Athey, 2017a; Varian, 2014; Belloni et al, 2014; Brynjolfsson & McAfee, 2014)。

从理论上说,机器学习(尤其是深度学习)是人工智能的一个组成部分(Taddy, 2018),更是当前人工智能领域科研和应用的重要驱动力,其算法的快速发展将会引起市场决策和政策研究的根本性变革(Agrawal et al, 2018)。为了加快我国在人工智能领域的技术进步,国务院于 2017 年 7 月发布了《新一代人工智能发展规划》,旨在 2030 年使我国的人工智能理论、技术与应用在总体上达到世界领先水平,并成为世界主要的人工智能创新中心。需要说明的是,《新一代人工智能发展规划》在强调发展高级机器学习、类脑智能计算、量子智能计算等前沿理论的同时,还特别指出在研究过程中要积极推动人工智能与心理学、数学、经济学和社会学等相关学科的交叉融合,以及开展跨学科的探索性研究^①。2017 年 10 月,在科技部征集的重大变革性科研技术项目中,有 4 项与人工智能有关,其中一项就是针对下一代深度学习理论与技术,力求突破我国在机器学习领域理论基础薄弱、模型结构单一的瓶颈^②。由此看来,探讨机器学习对经济学研究方法的影响,既可以帮助我们更好地理解机器学习的内涵,又有利于我们更多地使用机器学习工具以改善经济学研究方法,并在提高研究精准度的同时不断拓展新的研究领域。

在本文中,我们试图将机器学习的方法融入经济分析的框架中,并从以下几个方面梳理了其对外国经济学研究的影响:(1)机器学习的定义以及分类;(2)在大数据的背景下,阐述机器学习从哪些维度丰富了经济数据;(3)介绍机器学习如何改进了经济预测的准确性;(4)机器学习在因果推断中的应用。

^{*} 黄乃静、于明哲,中央财经大学经济学院,邮政编码:100081,电子邮箱:huang.naijing@gmail.com; mingzhe1619@163.com。感谢匿名审稿人的修改意见,文责自负。

一、机器学习的定义

机器学习不仅提供了新的分析工具,而且试图去解决传统计量方法所不能解决的问题。机器学习大致可分成四类:有监督机器学习、无监督机器学习、强化机器学习和半监督机器学习。

1. 有监督机器学习。有监督机器学习(Supervised machine learning)主要围绕着预测问题展开:需要使用被标注过的样本数据,包括特征(x)和对应的标签值(y),通过学习特征与标签值之间的对应关系,举一反三地构建出一个最优模型(Kotsiantis et al,2007)。这样,当面对新的特征 x 时,我们就可以使用这个最优模型来预测它的标签值 y 。这种通过已知数据寻找最佳估计参数的过程,与计量经济学中的回归分析有着异曲同工之妙。但是,有监督机器学习并不需要提前设定 y 与 x 之间的关系,可以根据数据本身选择更灵活的函数形式,可以避免过度拟合的问题,并兼备很好的样本外预测能力,这又与传统经济实证分析中的思路很不一样。在基于经典计量模型的实证分析过程中,人们往往关注的是 x 变量前参数的估计和统计推断是否准确。而机器学习的算法则侧重于 x 变量对于 y 的预测效果,并不太在意参数的估计是否准确。这也是为什么机器学习所估计出来的参数往往不满足一致性。目前,常用的数据分析软件(如Python、R等)中都有相应的软件包,可以较为轻松地实现这类算法(Pedregosa et al,2011)。

2. 无监督机器学习。现实中经常会出现这样的情况:缺乏足够的先验知识,或者使用人工的方式对数据进行标注的成本太高,从而只能获得包含特征(x),而没有相对应标签值(y)的数据样本(Gentleman & Carey,2008)。由于没有标签值,我们就无法根据误差(预测值与标签值)来找到最优的预测模型。因此,无监督机器学习(Unsupervised machine learning)要解决的是另外一个问题:当面对没有被标注的视频、图像和文本类数据时,我们希望机器可以从庞大样本集合中发现特征之间的潜在规律,识别出一些代表性的特征,并把具有相似特征的个体聚合在一起。需要说明的是,机器在分类时几乎不加入人类的判断,从而最大限度地体现客观性。无监督机器学习中最典型的例子就是聚类(clustering)。Athey et al(2017)在研究谷歌公司关闭西班牙境内的谷歌新闻对其新闻报道类型的影响时,使用机器学习技术对新闻进行分类。在给数据集降维后,研究关闭事件对同一时间内所有不同类型新闻报道的影响。这类技术在生活中也有广泛的应用,如当下流行的一些识别类软件(形色、微软识花等),可以根据用户拍下的一张花的图片,准确识别出花的类别。此外,在进行传统的回归分析之前,我们也可以首先使用聚类算法检验变量之间的相关性,对数据进行一些预处理,以降低计算量(Hodeghatta & Nayak,2017)。

3. 强化机器学习。有时,我们无法立即判断出一个决策的好坏,而往往要通过观察一系列决策实施之后的结果,才能对其进行评估。例如,在下棋时,某一步棋的好坏通常要在五步或十步之后才能判断。因此,与前两种机器学习不同,强化机器学习(Reinforcement machine learning)并不只是依靠已知的、固定的数据进行学习,而是在不断变化的环境中,通过大量、多次的试错学习,寻找到产生最佳结果的路径(Taddy,2018)。这类技术经常被应用到人工智能领域,如战胜人类围棋高手的阿尔法狗(AlphaGO)、多功能机器人和自动驾驶等。

4. 半监督机器学习。随着机器学习的深入发展,也随之出现了一些混合型机器学习技术,比如将有监督机器学习和无监督机器学习相结合的半监督机器学习(Semi-supervised machine learning)。这一机器学习方式专门用来处理已知数据中一些有标签值而另一些无标签值的情况(Chappelle et al,2009)。常用的半监督机器学习包括协同训练(co-training)、转导支持向量机(transductive support vector machine)等。

二、机器学习与大数据

“大数据”的出现改变了我们对于数据的认识:首先,“大数据”的基本特征是越来越庞大的数据数量。也就是说,样本中的观测数量非常多,甚至有可能是全样本数据。巨大的数据量对数据存储

和计算技术都提出了更高的要求。Varian(2014)介绍了一些大型 IT 公司处理数据的方法和使用的软件,比如可处理百万条大型数据的 SQL 技术、进行数据清理的 Open Refine 和 Data Wrangler 等技术。其次,“大数据”是由众多变量构成的高维度数据。随着文本挖掘、图像识别等技术的进步,不仅数据的类型变得越来越丰富(Blumenstock, 2016),而且对于每一个观测个体,我们都可以从任一有需要的维度进行变量挖掘。然而,要想将“大数据”应用于经济学研究,最重要的就是找到有效处理海量、高维数据的方法。对此,机器学习中的无监督机器学习技术可以处理相对于标准估计方法来说维度过高的数据,以帮助经济学家从图像、文本、音频等非常规数据中提取具有经济意义的信息。下面,我们将对应用以上几类新型数据的相关研究进行介绍。

(一)机器学习与图像数据

近几十年来,人造卫星一直在拍摄地球的图像,且大部分卫星图像数据都可以划分出数以千亿计的像素,而机器学习技术可以对不同的像素及像素组合进行处理,从中提取出有经济意义的信息,再与研究者感兴趣的变量特征进行匹配,从而构建出研究变量的替代变量。在分析经济学问题时,除了容易获取之外,卫星图像还有三个明显的优势:(1)可以使用卫星图像提供的信息替代难以测量或存在很大测量误差的变量。比如,由于印度尼西亚的林业受到高度管制,且当地官员有时会被贿赂以包庇非法采伐行为,导致官方统计数字通常是不准确的。Burgess et al(2012)研究印尼的森林砍伐问题时,根据卫星图像中细微的色彩变化,以更客观的方式测量了森林退化的程度,并以此估计了地区政治发展对自然资源的侵占。Jayachandran(2009)讨论了 1997 年因印尼森林大火所导致的空气污染对婴儿死亡率的影响。通过卫星传感器获取每日空气中的烟雾和灰尘数据,使得研究者可以持续地观察污染的整个传播过程。(2)卫星图像比传统数据具有更高的空间分辨率。Marx et al(2015)研究宗教信仰对市场投资行为的影响,由于贫民窟中的大多数房屋都使用铁皮屋顶,因此新更换的屋顶比旧的、生锈的屋顶反光率更高,使用 0.5 米分辨率的卫星图像就可以清晰地看到这种差异反射率,并以此衡量该家庭的住宅投资行为。当然,高空间分辨率也是相对的标准。Bleakley & Lin(2012)研究了地理环境对早期经济活动的影响,使用 1 公里分辨率的夜间照明数据,就可以在县级层面上衡量不同地区的商业活动水平,而一般的经济统计数据往往无法捕捉到上述差异。(3)卫星图像中包含地域广泛且时间跨度较长的高频率信息。Donaldson & Hornbeck(2016)研究气候变化对农业产出的影响时构建了一种基于卫星遥感数据的农学模型,这一模型就可以通过卫星数据估计世界范围内任何地区、任何农作物的产量的变化。

这种构建替代变量的思路特别适合讨论发展中国家的经济问题,由于发展中国家普遍缺乏关于经济发展的可靠数据,因此机器学习不仅解决了经济学研究中许多获取研究数据的问题,还可以用更贴近实际的数据不断丰富问题的研究角度。例如, Jean et al(2016)利用来自五个非洲国家(尼日利亚、坦桑尼亚、乌干达、马拉维和卢旺达)的高分辨率卫星图像数据,展示了如何使用机器学习技术来识别图像特征,并在高达 75% 的区域上研究经济产出的变化。这种方法的另一个优点在于仅需使用公开的数据,因此具有很强的扩展性,与依赖专业数据的其他方法相比,公开数据可以几乎无成本地对由于调查覆盖面较小而导致的测量误差进行校准,并在多国之间进行经济发展状况的对比分析。

除卫星数据外,一些学者还使用谷歌街道图像数据(Google Street View)研究社区内的房屋价格和居民收入问题。谷歌街景已经拍摄了 100 多个国家的街道和建筑环境,几乎包括了世界上所有主要城市的高清图像。Glaeser et al(2018)选择 2007—2014 年间 12200 张纽约和 3608 张波士顿的街景图像,利用机器学习技术中的视觉算法(Naik et al, 2015),识别出图像中建筑的物理属性(高度、距离及维护情况),并与美国社区调查(ACS)数据相结合,研究纽约、波士顿地区的房价和贫富差距问题。

(二)机器学习与文本数据

类似图像数据,机器学习也使得文本信息成了新的数据来源,从而构建出更具真实意义的经济指标。例如,在经历了近年来的全球金融危机、欧元区系列危机和美国党派政策纠纷后,人们普遍开

始担忧经济表现或经济政策的不确定性(economic policy uncertainty)所带来的影响。美国联邦公开市场委员会(Stekler & Symington, 2016)和国际货币基金组织(IMF, 2012)^③均指出, 美国和欧洲的财政、监管和货币政策的不确定性也是导致 2008—2009 年经济出现大幅度下滑的原因。他们认为严重的不确定性政策会引起企业和家庭缩减或推迟投资和招聘行为, 减缓经济复苏的步伐(Bloom, 2009)。

为了量化不确定性对地区经济发展的影响, 我们需要构建能够反映政策不确定性的指标。考虑到许多不确定性政策都是反经济周期的(counter-cyclical), 并且在经济状况欠佳时, 不确定性对经济的影响会进一步加强, 因此一般的做法是以手动的方式在新闻中挑选出与金融、经济危机有关的词语, 根据这些关键词出现的频率来衡量政策不确定性的程度(Alexopoulos & Cohen, 2009; Jurado et al, 2015)。然而现实中的情况是, 大多数文献通常使用一系列模糊的关键词, 如“经济”、“不确定”、“国会”以及“白宫”等, 这些词语并不能识别出新闻中不确定性的类型。尤其是对于那些几乎包括了所有关键词的新闻, 这使得我们很难研究不同类型的不确定性对经济发展的影响。例如, 一些与经济和金融危机有关的不确定性冲击会对企业的投资和招聘计划(Bernanke, 1983; McDonald & Siegel, 1986)、家庭消费(Fernández-Villaverde et al, 2011; Basu & Bundick, 2017; Leduc & Liu, 2016)以及金融市场的融资成本(Kraft et al, 2018; Gilchrist et al, 2014)等产生负面影响。但是, 另一些与技术和企业扩张有关的不确定性冲击会对经济增长产生积极影响, 且投资意愿会随着经济增长而加强(Segal et al, 2015)。

基于上述问题的新的解决方案是, 并不依靠某一组特定的关键词, 而是使用新闻中出现的所有词汇来估计新闻的主题, 从而将不确定性分为宏观经济的不确定性、经济政策的不确定性、金融政策的不确定性等多种类型, 这一区分方式可以分析不同类型的不确定性如何对投资和总产出产生影响。Larsen(2017)在对挪威最大的商业报纸《今日商业》(Dagens Naringsliv)在 28 年中的近 50 万条新闻进行分类时, 使用了机器学习中的隐含狄利克雷分配法(Latent Dirichlet Allocation, LDA), 又称“主题模型”。LDA 是一种无监督机器学习方法, 它假设每个新闻都是通过选择一个或多个主题, 并从这些主题中提取单词来构建的。如果令 T 表示主题的数量, 则一个单词出现在某一个新闻中的概率可表示为:

$$P(w_i) = \sum_{j=1}^T P(w_i | z_i = j) P(z_i = j)$$

其中, w_i 表示单词 i , z_i 表示单词 i 所来自的主题, 故 $P(w_i | z_i = j)$ 表示单词 i 来自主题 j 的概率。 $P(z_i = j)$ 表示新闻从 j 主题中选择单词的概率, 而每一条新闻都会从多个主题中选择不同的单词来构建。令 W 表示单词的数量, 则一个单词在各个主题中出现的概率可表示为:

$$P(w_i | z = j) = \phi_{w_i}^{(j)}$$

其中, 对于每一个 j 和 w_i , 都有 $j \in [1, T]$, $w_i \in \{w_1, w_2, \dots, w_w\}$ 。从而, 一个主题在一个新闻中出现的概率可表示为:

$$P(z = j) = \theta_j^{(d_i)}$$

其中, 对于每一个 j 和 d_i , 都有 $j \in [1, T]$, $d_i \in \{d_1, d_2, \dots, d_D\}$, D 表示新闻的数量。

根据以上算法, 作者选取了 1988—2014 年的新闻进行学习, 从中提炼出了 80 个主题和 150 个单词, 估算出每个单词在各个主题中出现的概率($P(w_i | z = j)$)。以图 1 中的两个主题为例, 单词的大小表示此单词出现在这个主题中的概率, 单词越大意味着出现的概率越高。接下来, 对于 2015—2016 年的新闻, 根据新闻中每个单词出现的频率, 使用学习后的模型逆向推断出新闻可能的主题结构, 以此来衡量不同新闻所表达的不确定性的类别及程度。

一些学者还使用这种方法构建了其他类型的经济指标。Larsen & Thorsrud(2015)创建了一个基于主题的新闻指数, 该指数用于研究新闻报道对挪威商业周期的影响。Hansen et al(2017)使用

主题1 宏观经济



主题2 货币政策



图1 两个主题分布的例子

资料来源:Larsen(2017),更多其他主题分布可参见 http://www.vegardlarsen.com/Word_clouds/。

主题模型对美联储发布的信息数据进行分类,研究增强信息透明度对货币政策制定者的影响。Hoberg & Phillips(2016)从美国上市公司关于业务描述的文本中总结公司的相似性,为这些公司生成了随时间变化的新行业分类。Kang et al(2013)利用点评网站上消费者对餐厅的文字评论与公共卫生部门的卫生检查记录相结合,研究了公开披露信息是否会影响餐厅的卫生状况。

(三)机器学习与数据集匹配

面板数据(panel data)是指在一段时间内追踪同一组个体的变化而形成的数据集,因此数据集中的信息来自横截面和时间的双重维度。由于面板数据及其分析方法具有控制异质性、降低多重共线性、减少数据偏倚性等诸多优点,因此,基于面板数据的计量分析方法及其应用主导了近年来社会科学界的经验研究(Wooldridge, 2010)。

虽然面板数据可以克服截面或时间序列数据的缺陷,但研究者往往很难获取样本容量充足的面板数据。对此,应用机器学习技术就可以把不同类型、不同来源的历史截面数据匹配在一起,提供更加完整的个人及家庭的面板数据集,以便分析早期生活状态、歧视和教育投资等问题的长期影响。就研究主题而言,经济史学家所面临的记录连接问题与现代数据集的问题截然不同。在现代数据集中可以使用唯一标识变量(如社保账号等)来连接不同的记录,但这些变量在历史数据中却极为罕见。相反,经济史学家们可以接触到其他的变量,这些变量可以被组合在一起以识别出某一个人,比如姓名、出生年份、出生日期和父母的出生地。但是,历史数据中的这些变量可能存在很多错误,包括转录错误、拼写错误、名称更改等等。Ferrie(1996)将1850年和1860年的联邦人口普查记录联系在一起,Thernstrom(2009)把美国两个城市中的几代人进行了配对,以研究代际流动问题。这些文章的做法都是雇用研究助理在不同数据集中搜索一个变量并进行手动匹配,但这种方法不仅耗费时间,而且由于不同研究助理判断匹配与否的标准存在偏差,也会造成结果的不一致。

Feigenbaum(2016)在研究收入的代际流动问题时,把1915年艾奥瓦州人口普查数据(X_1)中男孩的记录与1940年联邦人口普查数据(X_2)中成年男人的记录连接在一起。作者首先对 X_1 中的每一条记录,依据出生年份和姓名等信息,挑选出 X_2 中可能与之相匹配的记录。其中,出生年份的相似性通过年份的绝对差异来衡量,而姓名的相似性则通过Jaro-Winkler字符串距离来衡量(Goeken et al, 2011),由此可以得到一个出生年份差距小于3年,姓名距离小于0.2的数据集 XX 。接下来,作者在数据集 XX 中随机抽取一个子集 XX_T 作为训练集,然后人为地根据经验来判定 XX_T 中的每条连接 xx 是否是合格的匹配,进而从训练集中学习出一个最佳的算法,并以此挑选出数据集 XX 中所有合格的匹配。

三、机器学习与经济预测

经济预测是计量经济学的主要应用之一,准确的经济预测不论在宏观层面还是微观层面都有着十分重要的意义(Bernheim et al, 2013)。在前文介绍的几类机器学习算法中,有监督机器学习方法,

如回归树(regression trees)、LASSO、随机森林^④(random forest)、支持向量机(support vector machines)、神经网络(neural network)、深度学习(deep learning)等技术,因为其算法在预测上的优势,已被广泛地用于解决经济预测问题。

(一) 预测算法比较:有监督机器学习与传统方法

有监督机器学习算法的核心思路与计量经济学中的最小二乘法类似,其出发点依然是构建一个损失函数,基于该损失函数,通过代入数据来选择最优的预测模型。例如,当我们使用一个包含商品价格(y)和商品特征(x)的数据集(y_i, x_i)来预测数据集之外的商品的价格(\hat{y})时,我们首先定义一个损失函数 $L(\hat{y}, y)$ ^⑤,然后在数据集中找到一个具有最高拟合优度的函数形式 $\hat{y} = \hat{f}(x)$,使得预测结果与实际结果之间的平均损失达到最小值,即最小化 $E(L(\hat{f}(x), y))$ 。

然而,机器学习算法又不完全等同于传统的计量经济学方法。在一般的线性回归模型框架下,我们往往在回归方程的右边放入所有描述商品特征的解释变量 x_i ,使用数据集来估计每个解释变量的系数,从而对被解释变量进行预测。有时,为了改进预测能力,或者更好地解释某一解释变量的因果效应,我们会在回归模型中引入解释变量的交互项 $x_i x_j (i \neq j)$ 。但是,在解释变量数目很大的情况下,简单地把所有解释变量的交互项引入回归方程是不现实的,且可能会造成待估参数多于观测值个数的情况,并导致估计结果出现偏误,从而无法得到关于 y 的准确的预测结果。因此,选择哪些交互项放入回归模型直接关系到预测的准确度。例如,在预测人脸的模型中,需要使用大量、复杂的像素^⑥变量的交互项描述脸部器官的特征。在这种情形下,机器学习算法的优势就能体现出来,它可以自动地搜寻出符合预测要求的最优交互项组合。

以回归树分析法(regression trees)为例,Mullainathan & Spiess(2017)展示了机器学习方法在构建房地产价格预测模型上的优势。该文使用2011年美国国家户调查数据(American Housing Survey)中随机选取的10000个自住房数据进行分析,数据中主要变量为房子的价格,且每个房子都有150个与房屋特征有关的解释变量(如地点、面积、卧室数量等)。顾名思义,“回归树分析”法就是构建树形的预测模型,每个树枝的分叉处是树的一个节点,每个节点可以被看作是一个虚拟变量(即节点处有两条路径选择),虚拟变量的取值决定树枝分叉处的走向,如虚拟变量取1时向左边路径延伸,虚拟变量取0时向右边路径延伸。将树枝看作是一个线性函数,当到达树形末端(树叶)时,把所有的节点联结起来等同于将每个节点所对应的虚拟变量联乘,就可以得到关于房价的预测值。因此,回归树中的每一条路径都可以看作不同解释变量的交互项,整个回归树就是由多个交互项组成的函数族,树的深浅就代表着回归树的拟合程度。理论上,只要回归树有足够的深度,则每一片树叶只对应一个观察值,且每个观察值都可以由一组虚拟变量的交互项完美地拟合出来,此时就充分体现出了机器学习在算法上的灵活性。

(二) 机器学习预测方法的应用

一些学者已将机器学习方法应用于公共决策领域,预测的目的是提高社会生产率。例如,Krueger(2003)通过分析雇佣更多教师对学生成绩的影响,来决定学校的最优雇佣教师数量。Chalfin et al(2016)向我们展示了如何使用机器学习技术预测个人的劳动生产率,作者使用2009—2010年入职的664名数学教师和707名语言教师的调查数据,其中包括教师的知识水平、教学方式等信息、学生的考试分数、家庭环境等信息,以及学校教学评估的大量信息,基于机器学习方法估计教师任期决策的预测模型。随后,作者还使用美国费城警察局雇用的1949名警官的调查数据,其中包含丰富的警官个人信息和办公执法经历等,基于机器学习方法估计警察雇用决策的预测模型。作者发现,使用机器学习预测模型可以显著提高进行人事决策预测的准确性,进而改善社会福利。

此外,在一些司法管辖区,法官们每年都要做数百万次的判断来决定是否同意候审人的保释请求。为了解决法官的困难,Kleinberg et al(2017)使用2008—2013年间纽约市1460462个逮捕行动的案件数据,基于机器学习方法估计了候审人出庭率的预测模型。案件数据中具有丰富的信息,不仅包括在保释听证会上法官可以获得的信息(如候审人犯罪记录、所犯罪行等);还包括每个案件的

后续结果(如候审人是否被释放、是否在法庭上出席,或者是否在案件解决之前再次被逮捕等)。作者发现使用这种预测模型可以在保证预测准确度的同时,极大地缩短法官评判的时间,提高办理案件的效率。Björkegren & Grissen(2017)在构建信用评分体系时使用了机器学习方法,作者利用手机数据预测借款者的还款概率。由于许多发展中国家缺乏信用评分制度,许多个体也没有在银行或金融机构的交易数据。但是,研究者可以利用手机数据中与用户还款能力有关的行为特征来预测用户的还款概率。例如,一个负责任的借款者可能会将电话费保持在最低限度以上,以便他们在紧急情况下能够获得贷款,而一个容易出现违约的人可能会放任电话费用完,并依靠其他人给他们打电话。或者,一个通话次数很多的人可能拥有更强大的社交关系,意味着他们拥有更高的创业成功的概率。或者,费用支出的周期性可能反映了用户的不同职业,如工人通常在月初交费,商贩通常在市集日交费。作者最终从手机数据中提取了约 5500 个行为指标,并使用回归树方法来挑选最佳的预测模型。

机器学习预测模型也可以应用于个人决策问题。Kleinberg et al(2015)使用机器学习方法研究了老年人进行骨关节手术的决策问题。众所周知,骨关节炎(关节疼痛和僵硬)是一种老年人常见的慢性疾病,治疗的方法是通过手术更换髋关节或膝关节。然而这种手术不仅费用高昂,而且手术过程也很痛苦,且术后需要很长时间来恢复。如果进行手术的好处会随着生存时间的推移而增加,那么手术的回报率取决于患者术后的死亡率。文章使用 2010 年美国申请髋关节或膝关节置换手术的 65395 名患者组成的样本数据,包括患者的个人信息(年龄、性别、身体状况)和治疗信息(症状、伤害情况及其随时间的变化)。作者基于机器学习方法预测患者术后死亡率,如果预测结果显示患者将在一年内死亡,那么就不应选择进行手术。在一些关于行为经济学的研究中也出现了机器学习的身影。Camerer et al(2015)在研究动态的“非结构化谈判”(unstructured bargaining)问题时,使用机器学习方法来处理大量的谈判过程数据,以此来分析影响谈判结果的行为要素。Peysakhovich & Naecker(2017)研究了人们在面对金融市场风险时的选择行为,使用 600 名参与者的 6000 次选择数据,包括 55000 个与选择行为有关的特征变量及其交互项。文章对基于机器学习的预测模型和行为经济学中的标准预测模型(预期效用模型、非线性加权预期效用模型等)的结果进行比较后发现,基于机器学习的预测模型具有更准确的预测效果。

一些学者则把机器学习引入城市经济学研究之中,以解决城市资源的分配问题。例如,政府通常会组织专门人员去检查餐厅的卫生状况,对此,Glaeser et al(2018)使用机器学习模型来预测某个餐厅发生违规行为的概率,作者认为使用这种方法来指导卫生检查可以更合理地分配检查员负责的区域,在使检查频率降低 40% 的同时,不会减少发现违规事件的次数,这一结果显然提高了政府部门的工作效率。Goel et al(2016)使用机器学习方法来评估纽约市警察在“拦截盘查”政策的实施过程中是否存在种族歧视现象。他们使用纽约市 4 年中的 29 万次警察拦截事件,通过分析每条记录中 23 项与拦截事件、嫌疑人特征有关的变量,使用机器学习方法来估计个人持有武器概率的预测模型,作者发现,被拦截盘查的人中,黑人持有武器的概率相对于白人更低。

此外,机器学习预测方法也被大量应用在金融领域。Gu et al(2018)使用美国 3 万多只股票在 1957—2016 年间的的历史数据来预测这些股票的未来收益,且每只股票都有 94 个与股票收益特征有关的解释变量。在对股票收益进行预测时,典型的做法是将股票收益看成是一系列股票收益特征的函数(Fama & French, 1993; Lewellen, 2014),并使用包含股票特征的滞后项和少量宏观经济预测变量的回归方程来预测某一只股票的未来收益,但这种传统方法显然无法处理近百个变量在 60 多年中的信息。因此,作者基于 94 个股票收益特征、74 个行业虚拟变量以及 8 个宏观经济指标变量(如股息价格比率、国库券利率等)的数据,使用回归树、随机森林及神经网络等五种常见的有监督机器学习方法分别构建了股票收益预测模型。作者发现,在预测具有高流动性的股票或投资组合的未来收益时,所有机器学习方法都比传统的预测方法更加有效,且在五种机器学习方法中,回归树和神经网络的预测效果最好。同时,Rapach et al(2013)基于 LASSO 法利用世界主要国家股票收益的滞后项来

预测全球股票市场的回报。Sirignano et al(2018)使用神经网络法预测个人抵押贷款提前还款、拖欠和止赎的概率。

关于机器学习在预测中的应用,正如 Glaeser et al(2018)所言,这种预测方法可以在一个很微观的层面上比较不同的结果,从而评估各种政策及变化的影响。在未来的研究中,大规模成像和传感器技术的发展将会继续加深机器学习在生产力、福利分析领域中的应用。

(三)机器学习预测方法存在的问题

相较于传统的计量方法,虽然机器学习方法在预测方面具有一定的优势,但在政策分析中,机器学习方法可能存在如下问题(Athey,2017a)。

1. 机器学习模型是否具有有良好的样本外预测能力。尽管基于机器学习产生的预测模型在样本内(in-sample)具有完美的预测效果,但人们若使用该模型去预测样本外的个体往往会出现过度拟合(over-fit)问题,导致有偏的样本外(out-of-sample)预测效果(Yeomans et al,2016)。解决这一问题通常的做法是对预测函数的形式施加一定的限制(regularization)(Carrasco,2012)。以回归树为例,如果我们放弃追求树的“深度”(depth),则每一颗回归树的末端所对应的不再是一个观测值,而是一组观测值,这意味着在选择预测函数时允许适当的“噪音”(noise)存在。这样做必然会导致我们的预测树模型在样本内的预测效果变差,但却能改进我们的样本外预测效果。那么,我们应该对预测函数的形式添加多大程度的限制呢?或者说我们应该对预测函数的形式限制到怎样的程度呢?机器学习一般采用“实证调优”(empirical tuning)的方法,本质上就是把所有的观测值随机地划分成两类:一类被当作样本内观测值,用于估计预测模型,可以施加不同程度的限制;另一类则被看作是样本外观测值,用来检验这些预测函数的样本外预测效果,从中选取效果最好的预测函数。

2. 使用机器学习模型进行政策分析时可能导致不公平现象。Kleinberg et al(2015)对预测算法中的公平问题进行了深入的分析,作者发现使用机器学习模型对企业招聘、司法决策或借贷等行为进行预测时,存在性别、种族的歧视现象。但是,Athey(2017b)则认为机器学习能够促使社会资源分配变得更加公平。她认为,与人类相比,机器学习算法能够有效地吸收和利用更多的信息,并可以通过训练,在不同的约束条件下实现最大化目标。如果对一些算法生成的分配结果感到不满意,则可以通过调整约束条件来改进算法。

3. 机器学习预测模型的稳定性有待进一步验证。所谓的稳定性,是指可以保证在一个环境中训练出的模型到另一个环境中也能良好地运行。在机器学习中有许多相关的考虑,包括领域适应、转移学习等。值得注意的是,机器学习通过对 X 和 Y 之间可能存在的所有关联进行搜索,以寻找最佳的模型来预测 Y 。但是, X 和 Y 之间的某些联系可能在不同的时间和空间中发生变化。例如,在过去几年中,现象 A 和现象 B 总是同时出现,可以通过观察到的 A 来预测 B 。然而, A 并不是 B 不可分割的一部分,它们能够在不同的环境中独立存在。因此,在那些只存在 A 和 C 的地方,这种机器学习算法的性能就会受到严重的影响。对于一家每天都使用最新数据估计预测模型的科技公司来说,这可能不是问题。但在有些领域,预测模型的使用时期很长,甚至是固定的,这种情况下就需要寻找有效的方法来提高模型的稳定性。

4. 机器学习预测模型可能存在人为操纵的现象。Björkegren & Grissen(2017)指出,在一个使用手机数据预测借款者信用评分的模型中,借款者可能会对贷款提供者感兴趣的一些特征变量进行人为操纵。也就是说,如果某些行为模式可以帮助借款者获得贷款,那么借款者就可以编造出自己拥有这些行为模式的假象,欺骗贷款提供者。类似地,如果资源被分配给那些在卫星图像中看起来生存条件很差的家庭,那么家庭或村庄就可能通过改变他们房屋的空中外观,编造出自己很贫困的假象。Athey(2017b)指出,如何改进机器学习模型使其不容易被人为操纵,将是该领域未来研究的方向之一。

四、机器学习与因果推断

因果推断是计量经济学中的另外一个重要议题,其研究目的与经济预测存在一定的差异性。例

如,使用一个包含酒店价格和入住率的数据集,如果想要建立一个可以根据价格及其他因素(如地点、事件和天气等)对入住率进行估算的模型,这是一个预测问题。但如果想要研究一家酒店在价格上调后,其入住率将会如何变化,这就属于因果推断问题。前文所述的基于机器学习的预测模型,其研究目标与追求因果推断的经济学实证研究有一定的差异,但是两者并不矛盾,有相当一部分计量经济学家已经开始尝试将机器学习和因果推断结合起来,利用机器学习算法的优点来解决因果推断问题(Belloni et al, 2014; Chernozhukov et al, 2015; Komarova et al, 2015; Athey et al, 2016; Chernozhukov et al, 2017)。在本部分中,我们将对机器学习在因果推断中的应用做一个简单的梳理。

(一)估计平均处理效应

许多因果推断的文献都是在条件独立假设(conditional independence assumption, CIA)下估计平均处理效应(average treatment effects)。CIA 要求在控制了一系列协变量之后,潜在结果变量(potential outcome)独立于处理分配(treatment assignment),在此假设下计算处理组(treatment group)和控制组(control group)之间的平均差异,就是处理分配的因果效应(Angrist & Pischke, 2008)。但是,在实际研究中,往往存在大量的与处理分配和结果相关的协变量,人们并不知道哪些协变量是重要的。如果研究者没有控制住那些重要的协变量,就会导致估计的处理效应(treatment effects)是有偏的。Belloni et al(2014)和 Chernozhukov et al(2015)使用有监督机器学习中的 LASSO 法来选择影响潜在结果的最重要的协变量。传统最小二乘法通过最小化残差平方和来估计目标函数的参数,而 LASSO 法则在最小化残差平方和的计算中又加入了一个收缩惩罚项对估计参数进行缩减,即:

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j) + \lambda \sum_{j=1}^P |\beta_j| \right\}$$

其中, y_i 是潜在结果变量, x_{i1}, \dots, x_{iP} 包含了一系列用于控制的协变量(covariates)。显然,随着 λ 值的增加,一些重复的、不重要的协变量的系数会被缩减为零。同时可以在样本内随机选择同等大小的多个子样本,通过交叉验证法(cross validation)来选择最合适的 λ 值,并剔除此时系数为零的协变量。尽管 LASSO 的估计过程会带来处理变量系数的估计偏误(bias),但是我们却获得了更好的关于潜在结果的预测。为了满足条件独立假设,经济学家也会利用某一次外生(exogenous)的冲击,并使用工具变量法(instrument variable)来估计因果效应。机器学习方法也可以用在其第一个阶段的线性回归之中,能够提高工具变量对被解释变量的估计能力,从而缓解弱工具变量的问题(Mul-lainathan & Spiess, 2017)。

在评估某一个政策实施的因果效应时,我们经常使用双重差分法(difference in difference)。例如,想要研究新颁布的“营改增”税收政策对于私营企业投资行为的影响时,可以在政策试点地区随机抽取私营企业作为处理组,观察这些企业在政策实施前后投资行为的差别;同时,在与政策试点地区具有相似特征(地理、人口和经济发展水平等)的未实施政策的地区,随机抽取私营企业作为控制组,观察这些企业在政策实施时间点的前后投资行为的差别,两组差别再做差分就是税收政策对私营企业投资行为的因果效应。为了保证处理组和对照组的可比性,传统的解决方法有三个:(1)在回归方程中尽可能地加入反映地区特征的控制变量;(2)使用倾向得分匹配法(propensity score matching)挑选出与处理组得分相近的地区作为对照组(Abadie & Imbens, 2006; Diamond & Sekhon, 2013);(3)使用合成控制法(synthetic control)为处理组中的每一个单元(Unit)创造出一个与其特征极其相似但未被处理的单元,此单元是由控制组中的多个单元加权平均而得到的(Abadie & Imbens, 2011)。但是,这些方法通常无法处理存在大量协变量的情况。Doudchenko & Imbens(2016)将机器学习方法应用到双重差分的估计过程,不仅可以帮助筛选出重要的协变量。更重要的是,其与合成控制法的结合,能够为处理组的每个单元从控制组中挑选出一个最佳的单元组合,从而通过加权的方式创造出与处理组单元最相似的单元集合。进一步地,作者分别使用上述四种方式,重新估计了现有文献中三个使用双重差分法评估政策实施因果效应的经典案例——美国加州烟草法案

对吸烟率的影响(Abadie et al, 2010)、德国重新统一对西德地区经济发展的影响(Abadie et al, 2015)以及 20 世纪 80 年代古巴大量移民对美国工人工资的影响(Peri & Yassenov, 2015)。研究结果显示,使用机器学习方法能够更大程度地利用控制组中的单元信息,最大限度地满足处理组与控制组的相似性,从而可以显著地改善估计结果。

此外,断点回归法(regression discontinuity design)也是在估计因果效应时被广泛使用的一种方法。例如,在研究“新农保”政策对个人养老支出的影响时,由于存在身体素质等难以精确度量的其他因素也会影响养老支出,我们不能直接比较领保险和未领保险人群的平均支出差异。因此,必须要估计出那些领保险者如果未领保险时的支出,显然这种估计是反事实的(counterfactual)。“新农保”政策规定参保人在年满 60 周岁后可领取养老金,断点回归法就将 60 看作一个断点,并假设 59 岁未领保险者和恰好 60 岁可领保险者在身体素质上没有差别。因此,对于 59 岁者和 60 岁者来说,是否领取保险金就可以看作是随机的,59 岁者的养老支出就可以当作 60 岁者如果没有领取保险金时的反事实支出,从而 59 岁者和 60 岁者的平均支出差异就是“新农保”政策对个人养老支出的因果效应。但是,断点回归法对 59 岁者和 60 岁者具有相同身体素质的假设过强,1 岁的差别可能也体现着身体素质的差异,这依然会高估“新农保”政策的影响。机器学习为我们提供了另外一种估计反事实的方法(Varian, 2016)。可以使用小于和等于 59 岁者的样本,基于机器学习方法构建出未领保险人群中年龄与支出的模型,将 60 岁代入模型就可以预测出 60 岁但却没有领保险者的可能支出。理论上,利用机器学习方法在预测上的优势可以提高断点回归估计结果的精确性,但这一方法在实证操作中的效果仍然有待进一步论证。

(二)估计处理效应的异质性

在应用微观研究领域中,人们除了关心某一项政策实施的平均处理效应之外,还对其处理效应的异质性(heterogeneous treatment effects)感兴趣。比方说,我国现阶段正处于由人口资源大国向人力资源强国迈进的关键阶段,大力开展职业培训工作是实现这一转型的重要途径之一。政府部门在评估某项职业培训政策时,可能需要考虑该政策对于来自不同地区或者不同民族的人们,有着不一样的效果。在制定政策的过程中,充分考虑这种异质性的存在,能够提高政策在实施过程中的效率。传统的计量方法在研究处理效应的异质性时,依赖于在潜在结果回归模型中引入协变量与处理的变量的交互项。例如,我们想用以下回归模型研究某一项职业培训项目(用 Z 表示)对于未来工资(用 Y 表示)的因果效应:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \epsilon_i$$

其中 X_i 是控制变量(covariate), Z_i 表示第 i 个员工之前是否获得过业绩奖励(若获得过则等于 1, 否则为 0)。 γ 就是我们所关心的平均处理效应。但是,职业培训对于获得过业绩奖励的员工和没获得过的员工所带来的效果可能不同。我们可以在回归模型中引入 X_i 与 Z_i 的交互项,来研究处理效应的异质性:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \theta X_i Z_i + \epsilon_i$$

其中,交互项前的系数 θ 反映了职业培训的效果对于这两类员工的差异。

当协变量的个数比较少的时候,传统的计量方法(半参数方法或者非参数的方法)能够得到关于 γ 和 θ 的一致估计(Athey & Haile, 2007; Athey & Imbens, 2015)。但在面对大量协变量的情况下,这些方法可能会失效。这是因为,人们并不知道处理效应是随着哪些协变量而变化的。机器学习的方法可以在一定程度上改进异质性处理效应的研究(Asher et al, 2016)。例如, Athey & Imbens (2016)利用机器学习中的因果树方法,对整个协变量空间进行划分。划分协变量空间的目的在于找到最能体现出处理效应异质性的协变量交互项组合。这意味着,人们不需要预先设定处理效应是如何随着协变量而变化的,而是直接利用算法的优势去寻找答案。在划分后的协变量子空间内,人们可以计算受处理组和未受处理组之间的平均差异,即子空间内的平均处理效应。然后比较不同子空

间的平均处理效应,从而得到异质性处理效应的结果。进一步地,Athey & Wager(2017)引入随机森林法的概念,首先通过有放回的、随机的方式在样本集中抽取若干个相同大小的子样本,以建立一个包含许多因果树的因果森林(casual forest)。然后对于每一个因果树,从估计其处理效应的数据中找到协变量空间,使用 Athey & Imbens(2016)的方法进行计算,最后对众多因果树的结果取平均就是因果森林的结果。

(三)估计结构模型

不同于简约模型(Reduced form model),结构模型(Structural model)的构建是从经济体中的微观个体出发来描述主体的经济行为,从而使得研究者们能够分析经济政策的传导机制,例如宏观分析中常用的动态随机一般均衡模型(Dynamic Stochastic General Equilibrium Model)等。但是,大型的结构模型往往有很多的参数,研究者在使用传统的计量方法来估计结构模型时,会因为数据的限制得到有偏的估计(Bai & Ng,2008)。使用机器学习的算法,可以在一定程度上改进结构模型的估计效率(Abadie et al,2010)。

Ruiz et al(2017)使用从 5000 多件不同类别的商品中挑选出的信息来分析消费者的偏好,进而构建出可以揭示商品之间的替代或补充关系的模型。由于消费者数量和商品数量比较庞大,因此分析 5000 个产品之间的关系时就会有 2^{5000} 个组合,每个消费者的效用函数就会有 2^{5000} 个参数(即“效用—商品”的矩阵维度非常大)。并且在实际分析中,对于数目众多的商品来说,每个消费者通常仅购买极小的部分,因此矩阵也是极其稀疏的,很难用传统的方法进行估计。Ruiz et al(2017)使用机器学习中的矩阵分解法(Matrix factorization)来减少矩阵的维度,降维的核心思路是把“效用—商品”的矩阵分解成两个子矩阵。每个消费者通过购买商品获得效用时都有着不同的偏好,这些偏好可以用商品的特征来描述(如价格、形状、用途等)。由此,可以建立一个“效用—特征”的矩阵来表示消费者对这些特征所带来效用的偏好程度。同样,每一个商品也可以用这些特征来描述,由此可以建立一个“特征—商品”的矩阵来表示每一个商品中这些特征的含量。这样,两个子矩阵相乘就可以得到消费者对商品的效用。Wan et al(2017)使用类似的方法来构建消费者选择模型,其中包括消费者对商品类别、类别内商品购买数量的选择。然而,这种算法默认消费者会考虑所有 2^{5000} 个组合,并在其中进行优化,这在现实中不太可能实现。而且,消费者在做选择时往往是有限理性的,因为消费者通常会有顺序地向购物车添加商品。因此,将一些人类的自然约束融入结构模型中,最大限度地接近真实情况,可能是未来机器学习和经济学融合的研究方向之一。

另一项基于结构模型的研究来自 Athey et al(2018),几位作者利用美国旧金山海湾地区数千个手机用户的样本数据研究消费者选择午餐餐馆的行为。其中,餐馆具有一些可观察的特征,如餐馆的星级评定、食物类别及价格范围等。对于这些特征,每个消费者在选择餐馆时也有着不同的、个性化的偏好,即消费者对每个餐馆的光顾意愿和基本效用都是不同的。由此,作者基于矩阵分解法构建出描述消费者餐厅选择的模型。研究发现,相比于多标准的竞争模型(如多项 Logit、嵌套 Logit 模型),这种方法构建出的模型在预测消费者对餐馆开放和关闭时的反应方面表现得更好,并且可以分析当一家餐馆关闭后,消费者在其附近餐馆或拥有相似特征的更偏远的餐馆中如何重新分配他们的需求。最后,文章还展示了如何使用该模型来分析有关反事实的问题,比如在一个特定地点上,什么类型的餐厅能够吸引大多数消费者。

五、结论

在经济学研究中,理论驱动的推理和数据驱动的分析总是相辅相成、必不可少的两个部分。所谓数据驱动的分析模式,其目标就是让数据说话。机器学习提供了一个强大的分析工具,能够让研究者更清晰地听到数据所要表达的内容。未来,随着机器学习的发展,其对经济学研究的意义可能不仅仅是提供新的数据或新方法,而是要帮助经济学家发现新问题。本文对使用机器学习方法进行经济学研究的已有文献进行了较为系统的梳理,在与传统计量方法的对比中,阐述机器学习的优势。

首先,在面对大数据时,机器学习中的无监督机器学习技术可以处理相对于标准估计方法来说维度过高的数据,以帮助经济学家从图像、文本等非常规数据中提取出具有经济意义的信息。其次,在进行经济预测时,机器学习中的有监督机器学习技术可以根据数据选择更加灵活的函数形式,具有很好的样本外预测能力,同时又能避免过度拟合的问题。并且,在经济学实证研究最为关心的因果推断问题上,机器学习又可以通过挑选重要的协变量来改善平均处理效应的估计结果、估计处理效应的异质性,同时在估计结构模型时发挥重要的作用。

然而本文也指出,由于侧重点的不同,机器学习算法在经济学研究中所扮演的角色,更多的是对现有计量方法的补充,而非替代。机器学习侧重于预测效果的准确性,它可以利用灵活的函数形式挑选出一系列具有最佳预测效果的协变量,但却无法准确地估计出某一协变量的变化对于因变量的影响程度,以及相应的统计推断结果(Mullainathan & Spiess, 2017)。例如,在劳动经济学领域关于教育投入与回报的研究中,人们对增加受教育年限后,其工资是否会受到显著的影响,以及会变化多少非常感兴趣。机器学习的算法能够帮助我们找出一组变量,很好地预测工资水平,却无法准确回答受教育年限变量的显著性和影响程度的问题。而且,不同的预测变量组合也可能产生同样的预测效果,使得最终选择的预测模型可能没有很好的经济含义,这是机器学习算法的灵活性所带来的成本。再者,机器学习算法在实证调优的过程中,为了简化模型的需要,可能会遗漏重要的经济变量,导致我们关心的变量的系数估计产生偏误。在上述方面,传统的计量方法依然有其优势。因此,我们应该根据具体的问题,在研究过程中将两种方法结合起来。

注:

- ①国务院:《新一代人工智能发展规划》, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。
- ②中国科技部:《关于发布国家重点研发计划变革性技术关键科学问题重点专项 2017 年度项目申报指南的通知》, http://www.most.gov.cn/kjjh/xmsb/sbzj/201710/t20171010_135247.htm。
- ③国际货币基金组织(IMF), 2012, 49—53 <https://www.imf.org/external/pubs/ft/weo/2012/01/pdf/text.pdf>。
- ④随机森林是回归树的线性组合。
- ⑤损失函数 $L(\hat{y}, y)$ 可以是多种形式,最常见的形式是预测值和真实值之差的平方、绝对值等。
- ⑥像素(pixel)是指图像中被标明位置和色彩数值的小方格,而这些小方格的组合就决定该图像所呈现出来的样子。

参考文献:

- Abadie, A. & G. W. Imbens(2006), “Large sample properties of matching estimators for average treatment effects”, *Econometrica* 74(1):235—267.
- Abadie, A. & G. W. Imbens(2011), “Bias-corrected matching estimators for average treatment effects”, *Journal of Business & Economic Statistics* 29(1):1—11.
- Abadie, A. et al (2010), “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program”, *Journal of the American Statistical Association* 105(490):493—505.
- Abadie, A. et al (2015), “Comparative politics and the synthetic control method”, *American Journal of Political Science* 59(2):495—510.
- Agrawal, A. et al (2018), “Prediction, judgment, and complexity”, in: A. K. Agrawal et al(eds), *Economics of Artificial Intelligence*, University of Chicago Press.
- Alexopoulos, M. & J. Cohen(2009), “Uncertain times, uncertain measures”, University of Toronto Department of Economics Working Paper, No. 352.
- Alpaydin, E. (2014), *Introduction to Machine Learning*, MIT Press.
- Angrist, J. D. & J. S. Pischke(2008), *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press.
- Asher, S. et al (2016), “Classification trees for heterogeneous moment-based models”, NBER Working Paper, No. w22976.
- Athey, S. (2017a), “The impact of machine learning on economics”, in: A. K. Agrawal et al(eds), *Economics of Artificial Intelligence*, University of Chicago Press.
- Athey, S. (2017b), “Beyond prediction: Using big data for policy problems”, *Science* 355(6324):483—485.

- Athey, S. & P. A. Haile(2007), "Nonparametric approaches to auctions", in: E. Leamer & J. J. Heckman(eds), *Handbook of Econometrics*, Vol. 6, pp. 3847—3965, Amsterdam: Elsevier.
- Athey, S. & G. Imbens(2015), "A measure of robustness to misspecification", *American Economic Review* 105(5): 476—80.
- Athey, S. & G. Imbens(2016), "Recursive partitioning for heterogeneous causal effects", *Proceedings of the National Academy of Sciences* 113(27):7353—7360.
- Athey, S. & S. Wager(2017), "Efficient policy learning", arXiv preprint, arXiv:1702.02896.
- Athey, S. et al(2016), "Approximate residual balancing: De-biased inference of average treatment effects in high dimensions", arXiv preprint, arXiv:1604.07125.
- Athey, S. et al (2017), "The impact of aggregators on internet news consumption", Stanford University Graduate School of Business Research Paper, No. 17—8.
- Athey, S. et al (2018), "Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data", arXiv preprint arXiv:1801.07826.
- Baker, S. R. et al(2016), "Measuring economic policy uncertainty", *Quarterly Journal of Economics* 131(4):1593—1636.
- Bai, J. & S. Ng(2008), "Large dimensional factor analysis", *Foundations and Trends in Econometrics* 3(2):89—163.
- Basu, S. & B. Bundick(2017), "Uncertainty shocks in a model of effective demand", *Econometrica* 85(3):937—958.
- Belloni, A. et al(2014), "High-dimensional methods and inference on structural and treatment effects", *Journal of Economic Perspectives* 28(2):29—50.
- Bernanke, B. S. (1983), "Irreversibility, uncertainty and cyclical investment", *Quarterly Journal of Economics* 98 (1):85—106.
- Bernheim, B. D. et al(2013), "Non-choice evaluations predict behavioral responses to changes in economic conditions", NBER Working Paper, No. w19269.
- Björkegren, D. & D. Grissen(2017), "Behavior revealed in mobile phone usage predicts loan repayment", arXiv preprint, arXiv:1712.05840.
- Bleakley, H. & J. Lin(2012), "Portage and path dependence", *Quarterly Journal of Economics* 127(2):587—644.
- Bloom, N. (2009), "The impact of uncertainty shocks", *Econometrica* 77(3):623—685.
- Blumenstock, J. et al(2015), "Predicting poverty and wealth from mobile phone metadata", *Science* 350(6264):1073—1076.
- Blumenstock, J. E. (2016), "Fighting poverty with data", *Science* 353(6301):753—754.
- Brynjolfsson, E. & A. McAfee(2014), *The Second Machine Age*, New York: Norton.
- Burgess, R. et al(2012), "The political economy of deforestation in the tropics", *Quarterly Journal of Economics* 127(4):1707—1754.
- Camerer, C. et al(2015), "Dynamic unstructured bargaining with private information and deadlines: Theory and experiment", Caltech Working Paper.
- Carrasco, M. (2012), "A regularization approach to the many instruments problem", *Journal of Econometrics* 170 (2):383—398.
- Chalfin, A. et al(2016), "Productivity and selection of human capital with machine learning", *American Economic Review* 106(5):124—127.
- Chapelle, O. et al(2009), "Semi-supervised learning", *IEEE Transactions on Neural Networks* 20(3):542—542.
- Chernozhukov, V. et al(2015), "Valid post-selection and post-regularization inference: An elementary", arXiv preprint, arXiv:1501.03430v3.
- Chernozhukov, V. et al(2017), "Double/Debiased/Neyman machine learning of treatment effects", *American Economic Review* 107(5):261—265.
- Diamond, A. & J. S. Sekhon(2013), "Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies", *Review of Economics and Statistics* 95(3):932—945.
- Donaldson, D. & R. Hornbeck(2016), "Railroads and American economic growth: A 'market access' approach", *Quarterly Journal of Economics* 131(2):799—858.
- Doudchenko, N. & G. W. Imbens(2016), "Balancing, regression, difference-in-differences and synthetic control

- methods: A synthesis”, NBER Working Paper, No. w22791.
- Fama, E. F. & K. R. French(1993), “Common risk factors in the returns on stocks and bonds”, *Journal of Financial Economics* 33(1):3—56.
- Feigenbaum, J. J. (2016), “Automated census record linking: A machine learning approach”, <http://scholar.harvard.edu/jfeigenbaum/publications/automated-census-record-linking>.
- Fernández-Villaverde, J. et al(2011), “Risk matters: The real effects of volatility shocks”, *American Economic Review* 101(6):2530—2561.
- Ferrie, J. P. (1996), “A new sample of Americans linked from the 1850 public use sample of the Federal Census of population to the 1860 Federal Census manuscripts”, *Historical Methods* 29(4):141—156.
- Gentleman, R. & V. J. Carey(2008), *Unsupervised Machine Learning*, New York: Springer.
- Glaeser, E. L. et al(2016), “Crowdsourcing city government: Using tournaments to improve inspection accuracy”, *American Economic Review* 106(5):114—118.
- Glaeser, E. L. et al(2018), “Big data and big cities: The promises and limitations of improved measures of urban life”, *Economic Inquiry* 56(1):114—137.
- Gilchrist, S. et al(2014), “Uncertainty, financial frictions, and investment dynamics”, NBER Working Paper, No. w20038.
- Goel, S. et al(2016), “Precinct or prejudice? Understanding racial disparities in New York city’s stop-and-frisk policy”, *Annals of Applied Statistics* 10(1):365—394.
- Goeken, R. et al(2011), “New methods of census record linking”, *Historical Methods* 44(1):7—14.
- Gopalan, P. et al(2015), “Scalable recommendation with hierarchical Poisson factorization”, UAI Conference Working Paper, pp. 326—335.
- Gu, S. et al(2018), “Empirical asset pricing via machine learning”, Chicago Booth Research Paper, No. 18—04.
- Hansen, S. et al(2017), “Transparency and deliberation within the FOMC: A computational linguistics approach”, *Quarterly Journal of Economics* 133(2): 801—870.
- Hoberg, G. & G. Phillips(2016), “Text-based network industries and endogenous product differentiation”, *Journal of Political Economy* 124(5):1423—1465.
- Hodeghatta, U. R. & U. Nayak(2017), “Unsupervised machine learning”, in: U. R. Hodeghatta & U. Nayak(eds), *Business Analytics Using R-A Practical Approach*, Apress, Berkeley, CA.
- Jayachandran, S. (2009), “Air quality and early-life mortality: Evidence from Indonesia’s wildfires”, *Journal of Human Resources* 44(4):916—954.
- Jean, N. et al(2016), “Combining satellite imagery and machine learning to predict poverty”, *Science* 353(6301):790—794.
- Jurado, K. et al (2015), “Measuring uncertainty”, *American Economic Review* 105(3): 1177—1216.
- Kang, J. S. et al(2013), “Where not to eat? Improving public policy by predicting hygiene inspections using online reviews”, in: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.
- Kleinberg, J. et al(2015), “Prediction policy problems”, *American Economic Review* 105(5):491—495.
- Kleinberg, J. et al(2017), “Human decisions and machine predictions”, *Quarterly Journal of Economics* 133(1):237—293.
- Komarova, T. et al(2015), “Estimation of treatment effects from combined data: Identification versus data security”, in: A. Goldfarb et al(eds), *Economic Analysis of the Digital Economy*, University of Chicago Press.
- Kotsiantis, S. B. et al(2007), “Supervised machine learning: A review of classification techniques”, *Informatica* 31: 249—268.
- Kraft, H. et al(2018), “Growth options and firm valuation”, *European Financial Management* 24(2):209—238.
- Krueger, A. B. (2003), “Economic considerations and class size”, *Economic Journal* 113(485):34—63.
- Larsen, V. H. (2017), “Components of uncertainty”, Centre for Applied Macro-and Petroleum economics (CAMP), BI Norwegian Business School Working Paper, No. 4/2017.
- Larsen, V. H. & L. A. Thorsrud(2015), “The value of news”, BI Norwegian Business School Working Paper, No. 0034.
- Leduc, S. & Z. Liu(2016), “Uncertainty shocks are aggregate demand shocks”, *Journal of Monetary Economics* 82: 20—35.
- Lewellen, J. (2014), “The cross section of expected stock returns”, Tuck School of Business Working Paper, — 128 —

- No. 2511246.
- Marx, B. et al(2015), "There is no free house: Ethnic patronage in a Kenyan slum", IGC Conference Working Paper.
- McDonald, R. & D. Siegel(1986), "The value of waiting to invest", *Quarterly Journal of Economics* 101(4):707—727.
- Mullainathan, S. & J. Spiess(2017), "Machine learning: An applied econometric approach", *Journal of Economic Perspectives* 31(2):87—106.
- Naik, N. et al(2015), "Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major US cities", NBER Working Paper, No. w21620.
- Pedregosa, F. et al(2011), "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research* 12(10):2825—2830.
- Peri, G. & V. Yasenov(2015), "The labor market effects of a refugee wave: Applying the synthetic control method to the Mariel boatlift", NBER Working Paper, No. w21801.
- Peysakhovich, A. & J. Naecker(2017), "Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity", *Journal of Economic Behavior & Organization* 133:373—384.
- Rapach, D. E. et al (2013), "International stock return predictability: What is the role of the United States?", *Journal of Finance* 68(4):1633—1662.
- Ruiz, F. J. et al(2017), "Shopper: A probabilistic model of consumer choice with substitutes and complements", arXiv preprint, arXiv:1711.03560.
- Segal, G. et al(2015), "Good and bad uncertainty: Macroeconomic and financial market implications", *Journal of Financial Economics* 117(2):369—397.
- Sirignano, J. et al(2018), "Deep learning for mortgage risk", Available at SSRN; <https://ssrn.com/abstract=2799443>.
- Stekler, H. & H. Symington(2016), "Evaluating qualitative forecasts: The FOMC minutes, 2006—2010", *International Journal of Forecasting* 32(2):559—570.
- Taddy, M. (2018), "The technological elements of artificial intelligence", in: A. K. Agrawal et al(eds), *Economics of Artificial Intelligence*, University of Chicago Press.
- Thernstrom, S. (2009), *Poverty and Progress: Social Mobility in a Nineteenth Century City*, Harvard University Press.
- Wan, M. et al(2017), "Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs", in: *Proceedings of the 26th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee.
- Varian, H. R. (2014), "Big data: New tricks for econometrics", *Journal of Economic Perspectives* 28(2):3—28.
- Varian, H. R. (2016), "Causal inference in economics and marketing", *Proceedings of the National Academy of Sciences*, 113(27): 7310—7315.
- Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- Yeomans, M. et al(2016), *Making Sense of Recommendations*, Harvard University Press.

(责任编辑:刘新波)

(校对:刘洪愧)