

基于机器学习模型的审计应用：内涵、模式与风险

吴勇 方君 王尚纯 余洁

一、引言

美国注册反欺诈审查师协会(Association of Certified Fraud Examiners, ACFE)与世界著名的统计分析软件提供商SAS(Statistical analysis system, SAS)2020年联合开展的一项跨行业全球调查显示,约有25%的组织有意在未来两年内应用人工智能(Artificial Intelligence, AI)和机器学习(Machine learning, ML)来侦测欺诈行为,意味着AI和机器学习必将成为未来反欺诈解决方案的技术发展方向。机器学习是人工智能的一个重要子集,其基本思想是尝试训练机器,以便使其能够采取类似于人类学习的方式从大量复杂数据中进行学习,并能够预测未来。随着互联网、移动通讯、5G技术的兴起以及计算机处理速度和数据存储能力的提升,机器学习系统已普遍应用于图像识别、语音识别、产品、服务和内容的精准配对与推送、客户流失预测、信用卡欺诈监测和供应商风险管理等相关领域。例如,从网络搜索引擎到社群网络的内容过滤,从电子商务平台精准的目标广告推荐到抖音平台同类型视频内容的精准推送等,这些均是机器学习重要的商业应用。

机器学习未来有可能对所有行业产生颠覆性影响,会计和审计也不例外。机器学习将会极大提升会计工作的效率和效果,会计循环的多个步骤均可以实现自动化,尤其是当源文件不是结构化和机器可读的格式而需要人工处理时(例如手写或扫描图像),机器学习的优越性将进一步凸显。而且,应用机器学习的语音识别功能,当聊天机器人与客户交谈的同时,可以自动下订单,生成销售发票,并将数据传输到公司的会计信息系统。此外,某些传统的会计程序也可能会变得过时。例如,对库存的实物清点或对基础设施或其他资产的检查,可以由配备视觉识别系统的无人机自动完成。在机器学习的赋能下,该系统可以识别物体,检查质量和状况,测量一些关键的参数。通过这种方法,诸如“先进先出”和“后进先出”等库存管理方法可以低成本地运行和改进。

机器学习也已经成功应用于审计领域的多项任务中,如审查源文件、分析业务交易和事项、评估风险等。学者们还探讨了利用机器学习来预测欺诈(Bao et al., 2020)、破产(Gentry et al., 2002)、重大错报(E. Hunt, et al., 2020)和会计估计(Ding等, 2020)。德勤首席创新官

乔恩-拉斐尔(Jon Raphael)预计,机器学习将大大改变审计的方式,因为它能让审计人员在很大程度上“避免在速度和质量之间进行权衡”。机器学习算法不再主要依靠重复性抽样技术,而是可以审查公司的整体数据,审计人员能够以更加全域的视角来组织实施更有针对性的测试,及时辨识公司整体的异常情况。此外,机器学习算法可以从审计人员对具体项目的结论中“学习”,并将同样的逻辑应用于具有类似特征的其他项目。更为重要的是,长期以来,在实证研究演绎推理的范式下,归纳推理方法在会计和审计的主流研究中被低估,机器学习正在引起人们对归纳推理方法的重新认识,而且机器学习也是演绎推理类研究方案常选的技术方法和工具。例如, Li(2010)使用了一种贝叶斯机器学习算法,挖掘分析财务报告中管理层讨论和分析部分(Management discussion and analysis, MD&A)隐含的预测性、前瞻性信息内容。Brown、Crowley和Elliott(2020)使用机器学习技术,评估财务报表披露的主题内容在预测故意误报方面是否具有增量信息。基于算法和技术的视角,机器学习通过提供可靠的审计证据和促进决策机制改善,有助于提升审计效率和审计质量(Hsieh, et al., 2020; Zhang, Cho和Vasarhelyi, 2020)。然而,机器学习的实践应用也引发了人们对其潜在偏见和道德影响的关注,如何消除机器学习可能产生的偏见,治理机器学习可能产生的伦理风险,开展负责任的人工智能研究也是值得关注的研究议题。

本文在系统诠释机器学习本质内涵的基础上,剖析了机器学习引发的由演绎推理向归纳分析的研究范式转变,分析了机器学习在模式识别、自然语言处理、视觉信息和决策等多个领域的应用发展,总结梳理了机器学习模型审计应用的模式、面向审计全生命周期的机器学习模型系统应用以及国际“四大”机器学习系统的应用实践,围绕机器学习模型可能产生的偏见和伦理风险,辨析了机器学习模型审计应用所产生偏见的类型与潜在的风险和影响,并提出了相应的应对策略,以便更好地发挥机器学习的整体效能。

二、机器学习的内涵及其新发展

(一)机器学习的内涵界定

Arthur Samuel(1959)创造性地提出“机器学习”这个

概念,将其界定为“一个研究领域,它赋予计算机不需要明确编程就能学习的能力”。Mitchell (1997)为机器学习提供了一个供广泛参考的定义:“机器学习领域关注的是如何构建能随着经验自动改进的计算机程序的问题”。两者都坚持类似的观点,即机器学习的目的是使计算机自动(而不是由人类明确编程)从历史数据中学习模式和趋势,并且迭代改进其学习性能,可通过其预测精度来衡量。作为人工智能的一个子集,机器学习特别是监督式学习,部署算法基于样本数据(称之为“训练数据”)建立一个数学模型,并评估所建立的模型在独立数据(称之为“测试数据”)上的预测性能。机器学习的最终目标是确定哪种模型能够对未来的未见数据产生最准确和可重复的预测,以完成特定的任务(Mitchell, 2006)。

机器学习利用分析模型,使用迭代方法从分析数据中进行自动化和持续化的学习,当训练学习的数据足够多时,通过反馈行动不断修正学习结果,可以识别稳健的模式并进行预测。机器学习和传统统计分析在很多方面类似,但在执行机制上有所不同。统计分析是基于概率论和概率分布,而机器学习的目的是找到最能预测结果的数学方程组合。因此,机器学习非常适合于分类、线性回归和聚类分析等问题。机器学习方法大致可以分为两类:监督学习和无监督学习。监督学习算法使用有标签的例子,这意味着有已知输出结果的输入。监督学习适用于历史数据可以用来预测未来结果的情况,例如确定哪些客户最有可能拖欠债务。无监督学习适用于输出变量上没有标签的情况。系统不会被“告知”假设的答案是什么,而是自己找出数据模式。无监督学习包含不同的技术,可用于交易性数据(如聚类分析),如果将其作为风险评估过程的一部分,有助于发现以前未预见的风险。除此之外,还有半监督式学习,它包含了标记和未标记的两类输出数据的组合。

人工神经网络是未来人工智能和机器学习的重要组成部分。与人脑中的神经元类似,人工神经网络通过节点连接数据,系统可以被设置为形成简单或多层神经网络。深度学习(Deep learning)将机器的计算能力与神经网络中的连接模式相结合,以理解复杂的关(Hinton & Salakhutdinov, 2006)。深度学习的蓬勃发展主要是由于图像处理能力的提升以及计算机硬件成本的下降和机器学习算法效能的进步。目前通过深度学习的方法,Facebook 的人脸识别技术识别率已经达到 97.25%(Taigman et al.,

2014)。

(二)机器学习引发研究范式转变

传统的实证研究方法为了探究变量之间的关系,基于演绎推理的方法论指导,一般倾向于首先构建一个强有力的理论框架来指导假设的提出,然后在一个狭窄的数据域上对假设进行检验,在过去这需要一个漫长而费力的数据收集过程。随着人工智能、区块链、大数据、云计算、物联网、5G 技术等的深度应用以及程序算法效率的提升,极大地提升了现代计算和存储能力,潜在地改变了之前的研究范式,转向为一种更受数据驱动和实验的归纳推理方法,研究者利用海量数据之间的关联关系,归纳发展解释模型。现代数据分析能力允许采用多次近似的方法(通常是以百万次甚至亿次试验逐步近似,以达到更好的估计效果)来归纳,而不是推导出一个近似的解决方案。Tukey (1977)开发了“探索性数据分析”领域,该方法类似于运行数百个不同变量和时间段的回归方程,以选择最佳拟合模型。越来越便宜的计算能力、海量的数据存储和庞大的统计库为归纳推理过程提供了便利。图 1 总结了演绎推理法和归纳推理法两类研究范式的区别。归纳学习的研究和建模是机器学习的核心课题之一。机器学习使用了大量的算法,以监督模式(例如随机森林)或无监督模式(例如 k-means)检查数据池的共性和趋势,将特定过程的结果与相关数据拟合,有效辨识出海量数据背后的模式特征和规律认知,以便发现异常、预测趋势,将对舞弊侦测、风险评估以及持续审计和监控产生重要影响,这种数据驱动的决策也将会对引发会计和审计领域的研究范式实现由演绎推理向归纳分析的转变(Vasarhelyi, Alles, Kogan, 2004; Ding, Lev, Peng, Sun, Vasarhelyi, 2020)。

(三)机器学习的典型应用领域

数字时代,伴随着计算能力的提高、云计算和各类先进算法的推陈出新以及海量大数据资源之间的互动,这些均将极大地拓展机器学习在模式识别、自然语言处理、视觉信息和决策等多个领域的应用(麦肯锡, 2017; 郭

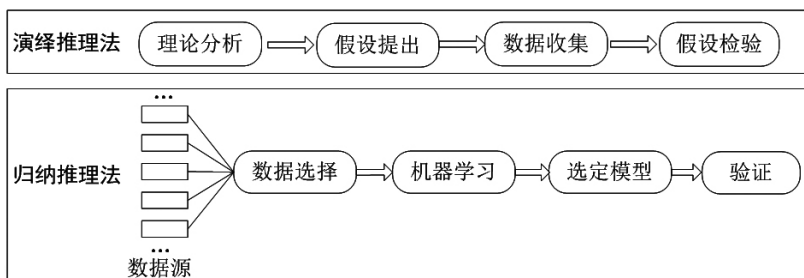


图1 演绎推理法与归纳推理法的比较

宇晨和张国宏,2020;赵军,2020)。

1. 模式识别。机器学习通过提取特征并相应地对数据进行分类/聚类,来识别数据中的模式(Bishop,2016)。为了获得更好的精度,往往需要更大的数据集。在大数据时代,各种数据源(包括来自社交媒体的人类生成数据和来自物联网的机器生成数据)提供了具有许多特征的高维信息。同时,带有内核函数的机器学习算法能够在更高维度的空间(即更多的数据列)中,对不易用多项式维度分类的数据进行分类,最终提高分类的准确性。不断增长的计算能力降低了计算成本,提高了计算速度,可以处理更多的数据(即更多的数据行),并通过覆盖、准备、整合和快速分析数据的实时过程,来支撑分类和聚类的准确性。

2. 自然语言处理。机器学习方法的一个热门应用是自然语言处理。自然语言类的文本数据相当一部分是由非结构化数据组成,机器学习可以通过按内容分类或聚类,从文本中提取信息。伴随着计算能力的提高,机器学习能够处理诸如语音、视频记录的转录以及来自社交媒体等各种类型实时文本数据。Sun(2019)利用机器学习方法,从财务报告电话会议的语句中提取反映情绪和情感的数据,辅助决策产生商业价值。

3. 视觉信息。机器学习可以通过视觉分析和数据可视化的方式来提升视觉信息。例如,图像和视频信息构成了互联网上最大的大片大数据,这一类数据信息由于其巨大的规模和复杂性,而目前的计算能力尚难以有效处理,产生了大数据应用的重要挑战。而深度学习是一种可以应用于搜索、检测和聚类视觉数据的机器学习技术(Hohman et al., 2019)。此外,机器学习允许整合文本数据等其他类型的数据,并通过数据可视化的过程中将其转换为可视化数据,通过可视化的数据,极大地提升了用户对数据的理解。例如,零售商停车场的可视化数据可以为其销售提供辅助信息。

4. 决策。机器学习的主要贡献,也是机器学习与其他算法的主要区别在于其预测能力,这种预测能力来自于训练和测试数据集的过程。具体而言,如果能够通过高效的数据预处理方法,将由自然语言和视觉信息组成的大数据清洗为合理干净的数据集,并具有较高的计算能力,那么机器学习算法就可以通过训练数据识别模式来更好地理解数据,以便提取关键信息,且能最小化人为干预对结果偏见的影响,提高预测精度和预测能力,最终能够协助管理者或决策者获得对其业务的更好洞察和见解,辅助他们做出科学决策(图2)。

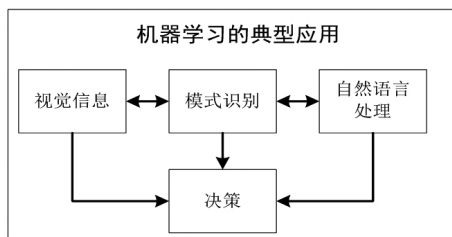


图2 机器学习的典型应用

三、基于机器学习模型的审计应用模式

(一)基于机器学习模型审计应用模式的运行机制

机器学习凭借可用的数据集和强大的计算能力和学习能力,助力自动化完成工作任务。基于机器学习的人工智能系统从训练数据集开始(如视频、文本、图像、声音等内外部海量数据源),通过应用数据分析和处理技术来检查交易的联系、模型、结构、合法性程度,并将学习算法应用到已经处理过的数据中,运行机器学习软件来从分析数据中识别模式,最终给那些包含潜在欺诈风险的交易贴上标签。隐藏在海量数据背后的模式识别过程是通过机器学习的迭代算法予以实现,机器学习算法使得机器的自主学习和动态适应成为可能。对于监督式学习,由于收集了某一输入的正确输出的例子,计算机系统可以“学习”并做出最准确的预测,以有效预测和识别潜在的欺诈、错误或可疑的交易。即使后续的数据发生变化,机器学习系统也会遵循同样的路径,通过在新的数据上训练机器学习算法(如识别与业务类型的不符的会计处理方式和会计估计方法,识别与原始会计凭证相比呈现出异常特征的交易和事项),系统能够对新的包含潜在欺诈风险的交易进行标注,可以为审计人员进行风险评估以及确定重点审计领域和审计范围提供重要的决策支持。如此,可以将大量的数据分析工作以及重复性的工作交由机器学习系统处理,不仅能够提升审计效率和质量,而且可以让审计人员有更多的时间进行审查和分析,这将使他们能够聚焦于高风险领域,以及更好地了解全局。机器学习的审计应用模式如图3所示。

(二)面向审计全生命周期的机器学习模型系统应用

机器学习的应用与实施可能会影响审计过程的所有阶段,以支持从审计计划到审计报告决策的全过程(Raphael,2017)。例如,在审计证据收集阶段,审计人员能够通过使用机器学习算法开发的规则收集数据,获得无偏见和更准确的信息。机器学习可以通过处理自然语言和视觉数据,提供来自不同来源的大数据的综合信

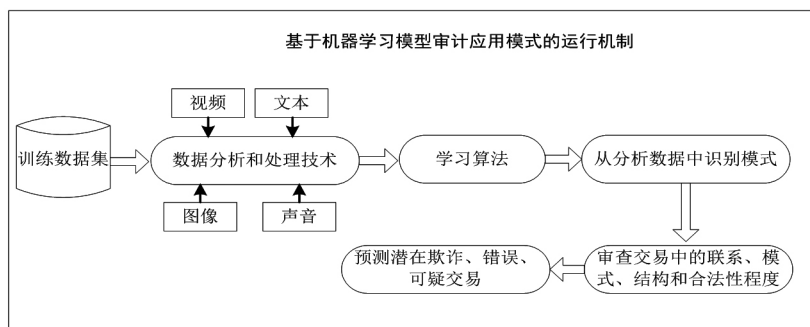


图3 基于机器学习模型审计应用模式的运行机制

息,审计人员可以在机器学习模式识别的帮助下,通过对这些额外的和更高质量的数据集的比对分析来识别异常值,还能够通过比较实际数据和机器学习生成的预测数据来检测异常。在风险评估阶段,语音和面部识别技术在鉴别欺诈访谈方面可以对审计人员起到重要的辅助补充作用。Ting Sun 和 Miklos Vasarhelyi 指出,在对高管进行舞弊访谈时,语音识别技术可以识别当受访者给出“有点”“也许”等模糊可疑回答时,可能暗示着欺骗,还可以识别出回答中的重大延迟也可能表明隐瞒。面部识别技术也可以应用于欺诈访谈,例如利用面部识别技术可以识别面试过程中暗示过度紧张或欺骗的面部模式。Sun(2019)提出了一个将深度学习应用于所有审计阶段的框架。该框架阐释了深度学习的文本理解、语音识别、视觉识别和结构化数据分析等能力如何适应审计环境,并且通过实例说明深度学习的信息识别和判断支持两个功能可以应用于多个审计流程,使得控制测试和实质性测试重要的审计工作均可以实现自动化。

此外,作为风险评估程序的一部分,审计人员可以通过使用机器学习工具,充分利用非传统的数据关系,以便更好地理解数字背后的驱动因素和行为模式,可以让审计人员侦测到目前可能被忽视的问题,如果发现与传统指标不一致的地方,恰恰揭示了审计人员需要关注的重点。例如,利用机器学习技术,可以描绘出一家餐厅停车场卫星图像相关的数据、从销售点系统获得的客人数量信息以及餐厅员工识别出高收入与高峰时段停车场的车辆数量信息等,通过辨识这些模式背后隐藏的数据内在关系,系统可以为审计人员所提供有价值的线索,识别并聚焦于销售收入与车辆数量、客人数量不一致的餐厅,而不是随机选择餐厅,如此可以极大地提升审计效率和审计质量。

(三)国际“四大”机器学习系统的实践应用

机器学习作为人工智能的一个重要子集,旨在尝试通过训练机器,以便使其能够采取类似于人类学习的方

式从大量复杂数据中进行学习,并能够预测未来。国际“四大”会计师事务所积极探索机器学习在审计业务中的应用,在审计计划、风险评估、交易测试、分析和编制审计工作底稿等审计实践中,机器学习正被用于执行审查总账、税务合规、编制工作文件、数据分析、费用合规、欺诈检测和决策等(Bowling&Meyer 2019;Kokina&Davenport,2017)。德勤(Deloitte)的 Argus 是一种机器学习工具,

Argus 学习编程算法使其可以读取租赁、

衍生品合同和销售合同等文件,并能够识别关键合同条款以及趋势和异常值,而审计师则专注于解释文件的关键特征,特别关注内在风险最高的合同,从而提高审计的效率和质量(Raphael,2017)。普华永道(PwC)基于机器学习技术开发的 Halo(Kokina 和 Davenport,2017),可以分析日记账并能够有效识别潜在的问题领域,例如带有可疑性质的关键词的条目,来自未经授权来源的条目,或异常多的日记账金额刚好接近授权的限额。Halo 允许审计人员测试公司一年内的每一笔分录。通过对所有分录进行测试,只关注风险最高的异常值,极大地提高了测试程序的效率和质量。安永(EY)成功开发了两款全球化的数字平台——EY Canvas 和 EY Helix,综合应用机器学习和自然语言处理技术,通过交互式视觉分析将调查人员与世界一流的机器学习无缝连接,旨在让用户降低法律风险并尽可能地收集调查需要的信息。交互式的视觉分析将非结构性数据中的关键内容用图表的形式进行呈现。调查人员可以通过改变图表的设置,快速定位与调查相关的文件。自然语言处理可以对所有文件进行分析并产生概念集群。调查人员通过标注关键文件,机器学习将对剩余文件与关键文件的关联性进行预判并给出预判分。调查人员可以根据文件的关联性预判分,有选择地对剩余文件进行审阅。毕马威(KPMG)建立了自己的人工智能工具组合 KPMG Ignite,旨在加强数字平台上得到业务决策和流程,包括多项人工智能工具。例如,呼叫中心分析引擎,利用自然语言处理设计模型来预测未来的事件,可以将客户来电转为非结构化文本,并将其精简找出关键字,用来预测客户的情绪;文件合规评估引擎,使用人工智能来阅读完整文件合同、租赁和投资协议,并产生重要相关信息。

四、机器学习模型审计应用的偏见和伦理风险

机器学习模型存在的一个问题是“过度拟合”(overfit)。数据科学中过度拟合模型被解释为一个从训练集(training set)中得到了高方差(variance)和低偏差(bias),

导致其在测试数据(testing set)中得到低泛化(generalization)的模型,也即计算机过度依赖于训练数据,在数据中发现了不代表现实世界中的特异性。过度拟合可能会导致机器“忘记”变量之间的统计学意义上的相关性,并不一定意味着因果关系。相反,“欠拟合”(underfit)发生在模型不够复杂,未能从训练数据中获得的信息,使其无法发现输入和输出数据之间的内在联系。对数据的过度关注会导致过度拟合,对数据的忽视又会导致欠拟合,找到最佳平衡点至关重要。因此,人类的理解和判断会在机器学习中发挥重要的作用。由于人类偏见的存在,可能会影响到训练数据集的选择、学习过程模型和方法的使用以及学习结果输出的解释等。因此人类需要对输入数据的性质,数据的学习和处理过程以及最终的输出有基本的认识。

(一)机器学习模型应用的偏见与伦理风险

Mitchell(1980)最早提出机器学习中的“偏见”(Bias)的概念,意指从经验中进行概括以预测新实例的标准。然而,今天机器学习中的“偏见”主要是指算法的任何“非预期或潜在的有害属性”(Suresh 和 Gutttag,2019)。例如,亚马逊的人工智能招聘系统存在性别偏见,在某些职位上更加偏向男性;Google Photos 和亚马逊的 Rekognition 在图像分类中存在性别和种族偏见;一款广泛使用的刑事累犯预测工具 COMPAS 被发现产生了带有种族偏见的结果,它对黑人被告的累犯预测显著高于白人被告。

对于企业而言,机器学习既是管理风险的工具,也是大量新风险的来源。因此,企业在采用机器学习模型、技术和工具时,应该识别各种机器学习技术内在的风险,以及业务部门在具体技术应用过程中所面临的风险,例如训练数据质量低下和编程问题,以及数据隐私法规和客户期望等外部压力。与机器学习技术相关的主要风险包括:(1)算法歧视风险。机器学习算法能够识别数据中的模式,并将其编写在预测、规则和决策中。如果这些

模式恰好反映了现有歧视状况,那么机器学习算法很可能会放大这种歧视,进而得出强化现有偏见模式的结果。(2)过高估计机器学习技术能力的风险。机器学习等技术受到热捧,因此其现有能力很容易被夸大。鉴于机器学习系统本身无法理解所执行的任务,而是依赖于训练数据,因此这种技术并非万无一失。如果机器学习系统的输入数据存在歧视、不够完整或质量欠佳,那么其结果的可靠性就会受到影响。(3)编程错误风险。任何计算机程序都无法避免出错,机器学习也是如此。当存在错误时,算法将无法按照预期运行并有可能得出误导性结果。(4)网络攻击风险。企图盗取个人数据或某公司机密信息的黑客可能会从目标公司的机器学习等智能系统下手。如果这些黑客控制了制定重要决策的算法,如驾驶汽车或控制机器人的算法,网络攻击的后果将不堪设想。(5)法律责任风险。目前,还鲜有管理机器学习和人工智能技术的立法,使得相关技术应用存在一定的法律责任风险。(6)信誉风险。机器学习等智能系统需要处理大量高度敏感数据,并负责在众多领域(包括信用、教育、就业和医疗)制定与个人息息相关的

表 1 基于机器学习模型的会计舞弊预测可能引发的不同类型偏见

偏见类型	偏见的内涵	面向会计舞弊预测的偏见示例	潜在的解决方案
历史偏见	当世界的现状与其历史地位之间出现错位时,就会产生历史偏见。	根据 SOX 之前的数据建立的会计舞弊预测模型,可能不适用于 SOX 之后的应用。	通过时间更新模型建立的训练数据。
代表性偏见	当训练过程中使用的数据未充分代表模型将被应用的对象时,就会出现代表性偏见。	在会计舞弊预测中,基于制造业企业组成的训练数据模型,可能不适用于服务业企业,因为不同行业的企业可能有不同的舞弊特征。	确保训练和测试数据代表模型将被应用的实际对象。
测量偏见	当所选择的特征或标签不能反映感兴趣的真实值时就会产生测量偏见,它们可能会遗漏重要因素或引入噪声,导致性能不同。	在会计舞弊预测中,用会计和审计报告上披露的舞弊事例来衡量实际舞弊行为。然而,所披露的欺诈行为可能遗漏了欺诈行为发生但未被监管机构发现的情况。	尽量使用准确的代理变量,并考虑不同群体的测量偏见。
聚合偏见	当根据汇总数据建立模型时,不同的人群被不适当地结合起来,就会产生汇总偏见。然而,由于各子组之间的异质性,单一的模式不可能适合所有子组。	根据所有行业的汇总数据建立的会计舞弊预测模型可能对某一行业的公司不适用。虽然聚合偏见与代表性偏见非常相似,但它更侧重于聚合数据的不当使用。相反,代表性偏见强调的是模型不是基于代表性的训练数据建立的错误。	将群体差异信息纳入模型中,以提高各群体的绩效。或使用分类数据对模型进行训练和测试。
评价偏见	当模型的性能没有使用适当的指标或有代表性的基准数据集进行评估时,就会出现评估偏见。	按准确度评估的会计舞弊预测模型不会有效,因为舞弊是罕见的小概率事件,而准确度可能非常高,即使该模型做出天真的预测,认为所有的情况都不是舞弊。	使用能代表将应用该算法的人群的基准数据集,并使用适当的评估指标。
部署偏见	当系统在实际采用过程中被以不恰当的方式使用或解释时,就会出现部署偏见。	会计舞弊预测工具是为了在财务报表审计过程中使用,但它只在审计前用于决定是否接受或继续审计工作。如果不按预期的方式部署工具,就无法获得模型的最大效益,也无法帮助改进和更新模型。	确保建立的模型是针对部署后实际执行的特定任务。

重大决策。因此,如果机器学习等智能系统存在歧视、容易出错、遭到黑客攻击或使用目的不道德,相关组织的信誉将会受到极大影响。

(二)机器学习模型审计应用偏见的来源和类型

偏见可能发生在机器学习的任何阶段。Suresh 和 Gutttag(2019)将机器学习项目的典型生命周期总结为六个阶段:数据收集、数据准备、模型开发、模型评估、模型后处理和模型部署。基于机器学习全生命周期的六个阶段,可以进一步将机器学习中的偏见分为历史偏见、表示偏见、测量偏见、聚合偏见、评估偏见和部署偏见。表1系统总结了基于机器学习模型审计应用过程中,不同类型偏见的内涵以及在预测会计舞弊领域的典型示例和潜在的解决方案(Cho 和 Vasarhelyi,2020)。

(三)机器学习模型偏见和伦理风险的应对策略

为了解决机器学习中的偏见问题,Munoko、Brown-Liburd 和 Vasarhelyi(2020)采用两个面向未来的伦理框架,研究人工智能方审计应用所面临的伦理问题。Appelbaum、Issa 和 Strauss(2020)提出了一个通用的框架,为审计人员在应用人工智能技术对财务报表信息审计时提供伦理遵循,以合理保证人工智能应用不存在潜在的偏见和不公平的结果。此外,新兴的公平机器学习领域提出了反分类(anti-classification)、分类均等(classification parity)和校准(calibration)等标准,以确保基于机器学习算法的决策是公平的(Corbett Davies 和 Goel,2018)。其中,反分类要求“诸如种族、性别等受保护的属性不被明确用于决策”;分类均等要求“在受保护属性所定义的组之间,预测性能的通用测度指标(如假阳性率和假阴性率)相等”;校准意味着“以风险估计为条件,结果独立于受保护的属性”。除此之外,建立简约、透明、可解释的机器学习模型也是解决偏见问题的重要路径(Bostrom 和 Yudkowsky, 2014)。会计和审计的从业人员以及相关监管部门也致力于建立透明、可解释的机器学习模型,开发准确、可解释的算法,一些商业工具可以用来检测和防止机器学习模型的偏见,如谷歌的 What-If Tool(WIT)和 IBM 的 AI Fairness 360。然而,关于人工智能算法的透明度和披露也可能导致意想不到的后果,例如算法被黑客攻击的脆弱性增加,以及更容易受到法律诉讼或监管行动的影响,从而产生“透明度悖论”问题(Burt,2020)。

五、未来研究方向

机器学习在会计和审计领域的应用已形成了丰富的成果,有力提升了审计效率、审计质量和审计价值,如下议题未来值得深入探讨:(1)尚需进一步加大机器学习在会计与审计领域研究的推广应用,以便证明机器学习在多种会计估计中的有效性,以提高财务报告质量。

例如,公司可以建立坏账准备会计估计的机器学习预测模型,利用公司和客户双方的交互数据进行训练,并将机器学习模型的预测结果与管理层给出的会计估计相结合,以便得到更加科学合理的估计。审计人员可以将机器学习模型提供的会计估计预测值作为一个判断基准,与管理层提供的估计值进行比较,以便确定需要进一步调查的领域。此外,准则制定者可以考虑按行业开发机器学习模型来预测不同类型的会计估计,并要求上市公司使用这些模型来生成会计估计。或者,通过机器学习计算出全行业的平均估计,并要求公司在其主要部门分别使用这些估计,这将有助于减少会计估计的主观性和人为偏见,但由此也将改变财务报告的范式,是一个值得探讨的课题。(2)可以深入探究会计和审计领域使用机器学习引发的偏见和伦理影响。虽然机器学习的偏见(尤其是种族和性别偏见)已经在贷款审批、犯罪预测和智能招聘等领域得到了广泛的讨论,但机器学习在会计和审计等特定情境的偏见尚未得到充分的研究,特别是当机器学习被用于会计和审计中的特定任务时,可能会发生哪些类型的偏见,以及如何解决这些偏见?面向会计与审计具体应用情境与决策问题,将机器学习产生的决策与人类的决策进行比较,以判定何种情形下会产生不同的决策结果,以及在什么情况下人类会推翻或怀疑机器学习的决策。与此相关的另一个问题是,考虑到人类也会有自己的偏见,那么基于一个不完美算法给出的决策,在某些情况下是否会比同样不完美的人类决策更值得信赖也值得研究?(3)学术界和政策制定者还需要考虑研究会计师和审计师使用的机器学习工具的可比性和能力要求。特别是,是否应该建立合理的标准(如对训练数据、测试数据、模型性能和模型更新的要求)来规范会计和审计人员对机器学习工具的使用?这些工具是否应该得到数据科学家的认证,并得到相关监管机构的批准?此外,还应该讨论在会计和审计决策中使用机器学习的潜在法律后果等。

基金项目:(1)安徽省教育厅教学研究项目省级重点项目“新一代信息技术条件下决策型会计人才培养模式研究”(2017jyxm0040);(2)国家自然科学基金专项项目“国家自然科学基金依托单位信用评价体系和管理机制研究”(J1924003)。

作者单位:合肥工业大学管理学院

原载《中国注册会计师》(京),2021.9.34~40

