

利用机器学习提升内部审计价值

——读《机器学习强化保证》

余玉苗 徐慧慧

电子信息技术的蓬勃发展掀起了人工智能的浪潮,机器学习(Machine Learning)作为人工智能的核心技术得到了更广泛的应用。在内部审计领域,经济业务越来越复杂,交易对象和交易环境也呈现出明显的多元化趋势,由此带来的新挑战层出不穷。机器学习可以为内部审计提供更智能的处理系统,从而提高审计质量和审计效率。随着机器学习技术的日益成熟,将机器学习引进内部审计领域的步伐也在不断加快。在这样的背景下,企业应该如何利用机器学习以实现内部审计领域的突破?新加坡政府投资公司的IT审计和数据分析负责人Ying-Choong Lee于2019年8月在《内部审计师》(《Internal Auditor》)杂志上发表了《机器学习强化保证:人工智能可以借鉴过去事例产生有用、真实审计洞见》(《Stronger assurance through machine learning: By inferring from past examples, artificial intelligence tools can generate useful, real-world audit insights》)一文,对人工智能时代下机器学习在内部审计领域的意义、应用和挑战做了阐述,值得企业在改进内部审计时参考。

一、何为机器学习

机器学习技术使计算机能够模拟人类的学习行为,从而获取新知识或新技能,甚至可以重新组织自身已有的知识结构并不断改善。这项技术是计算机智能化的根本途径,要想更恰当地利用机器学习,首先应对它有深入的了解:

其一,机器学习可以基于过去预测未来。这与传统的、基于规则的分析系统十分不同。虽然基于规则的系统也可以通过现有的计算机技术实现自动化,但是它们只有在提供了准确条件的情况下才能够处理输入。机器学习则“聪明”得多。它只需要考虑相关条件的输入,并将这些输入与过去的结果结合起来,就可以生成可能的预测。也就是说,对于机器学习而言,用来预测未来行为的最佳指标是过去的行为。

其二,机器学习对数据格外重视。这是因为机器学习算法结果的准确性在很大程度上取决于输入的训练数据集。在机器学习模型中,数据的特征和标签尤为重要。特征可视为模型中的“自变量”。为了从训练数据集中最大程度地提炼特征,需要配备特征工程(Feature Engineering),这是提高算法精度和效率的重要一步。也就是说,有效的特征工程可以减少训练数据集中特征的数量。这个

过程可以减少模型的假设和偏差,从而使其得到更好的推广。标签则是指需要被预测的事物,即可视为机器学习模型中的“因变量”。

其三,机器学习算法的学习方式通常可以分为有监督学习和无监督学习两大类。这两种学习方式可以适应不同的现实情况。有监督学习通过训练数据集的特征和标签进行学习。通过分析训练数据集,算法可以学习特征与过去结果之间的关系模型,然后将所学习的模型应用于新的数据集以获得预测结果。无监督学习则是一种不考虑过去结果的学习方式。这种学习方式根据特征的相似性、差异性或模式对数据集进行分组,在不考虑具体任务的情况下学习它们观察到的数据,从而创建自主智能。这种方式主要适用于标签数据昂贵或不可用的情况。

二、机器学习如何在内部审计领域应用

作者使用了支付交易的例子对机器学习在内部审计领域的应用进行了说明,主要体现在提高内部审计效率和进行更全面的分析两个方面。

一方面,机器学习可以提高内部审计的效率。这是因为机器学习对

特征执行多维分析的能力远超于基于规则的系统。尤其是当识别异常时,无论使用无监督学习抑或是有监督学习的模型都非常有效。无监督学习模型可以生成一组潜在的支付异常,并对这些异常进行跟踪,以确定它们是真是假。之后,识别的结果将被纳入机器学习模型,从而使模型可以用更高的精度识别异常支付。有监督学习模型则需要先训练一支历史支付错误算法的标签,用以提前识别潜在错误。除示例的支付异常外,这种模型还可以扩展到检测包括费用、采购和授权在内的其他各种环节的异常行为。

另一方面,机器学习可以在内部审计中进行更全面的分析。作者给出了四个示例:第一,通过观察每种产品的付款金额来确定服务是否存在超额付款的状况。并且,机器学习还可以突出显示所有明显偏离平均值的支付交易。第二,通过突出显示的支付异常值,识别出之前未知的、新出现的需求或客户类型,从而可以帮助高管优化产品以及降低服务成本。第三,判断是否有交易对手存在低于特定阈值的连续多次付款。这项分析将有助于识别出那些被分割成更小额度的可疑支付。第四,可以识别出向供应商或相关实体支付的重要款项。

三、机器学习在内部审计领域应用中面临的挑战

作为尚在发展中的新兴技术,机器学习在内部审计领域的应用也将面临许多挑战。Ying-Choong Lee 主要从以下三个方面进行分析。

一是机器学习需要大量高质量的、有标签的数据。经验表明,在机器学习的项目中,数据探索和数据准备通常消耗了最多的时间。如若提

供的数据与实际状况有所出入,那么将对输出的结果产生非常不利的影响。而如若没有配备良好的标签和特征工程,有监督学习模型的结果将发生偏差,可能会产生噪声或无意义的结果。

二是机器学习的算法模型可能存在一定程度的性能偏差。事实上,大多数利用机器学习的内部审计项目最初都可能会从一些模型中得到不准确的结果。潜在的原因可能包括:模型不能很好的泛化、特征工程有偏差、缺乏高质量的数据和使用了不适合底层数据的算法等。而过度拟合则是导致模型性能出现偏差的另一个潜在原因,这也是数据科学家经常遇到的一个状况。为了减少这种偏差,内部审计师可以使用独立于训练数据集的数据来验证模型的准确性。但同时,审计师也应该认识到每种算法的内在局限性。

三是目前机器学习的应用缺乏实践经验。主要体现在新采用机器学习技术的项目组没有成功的项目示例用以学习。作者认为,内部审计师可以先应用简单的机器学习模型,这有助于他们增强能力,提高信心。在拥有一定的经验后,再分阶段逐步实现对算法的改进。几乎所有新技术在最初应用时都会面临缺乏经验的困境,采用这种渐进的方式不失为一个有效的应对措施。

四、启示

笔者认为,Ying-Choong Lee 的这篇文章对机器学习进行了介绍,讲述了机器学习在内部审计领域的应用,也指出了机器学习在内部审计领域面临的挑战和未来的发展方向,对推动我国企业内部审计的智能化进程具有一定启示作用。

一是要加强对机器学习的了解

和重视。虽然我国目前对机器学习的研究和应用相对于国外而言还比较少,但我国的机器学习技术正在高速发展中。百度、腾讯、阿里巴巴三家互联网企业分别建立了PaddlePaddle、Angel、DTPAI三大机器学习开源平台,学术界也参与机器学习技术的研发浪潮,建立了SeetaFace,目前供学术界和工业界使用。随着大数据时代的发展,智能化已成为一个不可逆的趋势。企业应以积极的态度接纳机器学习,增强对此技术的了解和关注。

二是要强化对内部审计师相关知识的培训。机器学习技术的诞生对内部审计师提出了更高的要求。未来的内部审计工作越来越需要具有专业知识的审计师与人工智能合作完成。对于内部审计师而言,拥有一定的人工智能知识可以帮助他们更好地适应行业的变革。对于企业而言,积极地对内部审计师进行相关培训也有助于其面对机器学习带来的各项挑战。

三是要推动在内部审计领域恰当地应用机器学习技术。1983年国家审计制度的建立是我国内部审计制度发展的起点。虽然我国的市场机制在不断完善,经济也在迅猛发展,但是由于内部审计起步较晚,发展相对落后。而利用机器学习可以高效地获取知识,这将是内部审计行业智能化改革的关键,也是我国内部审计行业发展的契机。可以说,机器学习的出现给我国内部审计行业提供了一个弯道超车的机会。企业在完善内部审计时应重视这次机会,恰当地应用机器学习技术,以寻求更好的发展。

(作者单位:武汉大学经济与管理学院会计系)

责任编辑 李卓