

黑匣子里的 机器偏见

MACHINE BIAS INSIDE THE BLACK BOX

Brad J. Monterio
张环译

了解潜在的数据偏见以及怎样保护公共利益，应该是管理会计师未来的首要关注点。

数据分析存在偏见并不是什么新鲜事。随着人工智能、认知性机器学习、智能系统等更先进技术的兴起，以及其他一些基于数据发挥作用的项目不断发展，数据分析偏见进一步显现出来。一些数据专家认为数据永远不可能脱离偏见，是因为数据生命周期——从数据编写到收集、存储、验证、分析到报告或传递——都在某种程度上都与人类相关。我们难以避免产生偏见，即便是无意识的偏见。

一些开发人员认为，当大量数据在“黑匣子”内处理时，通常可以创建高级算法来规避偏见。但正如照片墙（Instagram）公司营销分析和决策科学负责人Chris Dowsett在《走向数据科学》（Towards Data Science）（bit.ly/2I9jGik）中提到的：“人类既是数据创造者又是数据使用者，这意味着偏见有机会渗入数据生命周期。”

换句话说，无论是编写数据还是在数据分析完成后对其进行诠释，人际互动都可能引入偏见，或者令那些用于驱动决策和经营策略的洞见变得难以理解，从而导致糟糕的预测和错误的决策。这对企业和受影响的个人来说，代价都非常高。

风险突显

作为IMA®（美国管理会计师协会）技术解决方案和实践委员会名誉主席以及IMA 多元化和包容性委员会现任主席，我的工作与这个问题密切相关，我将之称为“机器偏见”。我非常关注这一话题，它也应该引起你的注意。

我现在比以往任何时候都更有可能成为高级算法所带来的偏见的受害者，这些算法非常先进，能驱动应用高级分析工具



与人工智能来对我做出判断或预测，包括我未来的购买行为、工作前景或者保险保障范围。

我的一位同事在四大会计师事务所中的一家担任首席数据科学家，我们多次谈论过机器偏见的风险及其开辟的新领域。对于会计和财务专业人员来说，数据分析并不是一个陌生的领域，与之前相比，他们分析数据的环境已经发生了巨大变化。目前，越来越多的组织可以使用基于云的、可扩展的、经济高效的人工智能解决方案，而此前只有少数公司才能承担这类解决方案的成本。这种大众化的应用意味着管理会计师能够使用这些工具，并以近十年来前所未有的方式，对结构化和非结构化数据集进行预测性和规范性分析。

无论是大数据还是“小数据”（为个体客户、成员等提供的更个性化的数据），会计师都可以通过这些新技术对其进行审查、分析和可视化，以找到新的相关性、模式以及信息方面的洞见。对某些人来说，这些功能可以帮助他们学习与业务相关的全新知识技能，包括战略驱动因素等有益的东西；但它们也增加了潜在的风险，例如数据分析中的偏见或歧视。

无论是在当前还是更长远的未来，一家运作良好的企业将面临的难题之一，就是要更加重视并积极实践多元化和包容性（D&I）。而我们知道，人们会在数据生命周期引入偏见，这意味着组织可能会变得单一化和排他化。除此之外，还有高级



分析中使用的工具黑匣子，其嵌入在软件解决方案中的复杂算法，可以收集大量信息并得出某些结论，这会导致潜在的偏见被放大，尤其是在算法设计本身就存在偏见的情况下，尽管这种偏见可能是无意识的或无意的。

多年来，监管机构一直担心信息会成为企业对付某些利益相关者或个人群体的武器。例如，抵押贷款机构拒绝为高风险社区客户提供贷款，或者银行认为某个种族或族裔信用风险太高而不提供服务。在某些方面，事情并没有真正改变，甚至从未改变过。

监管困局

我们常常看到监管机构奋力追赶，希望在新技术方面与市场同步。例如，银行与资本市场监管机构刚开始部署企业部门早已使用的人工智能和其他新技术。问题在于，这里存在一个滞后阶段，在这期间，企业越来越成熟，技术更加精进，其负面效应也可能会急剧膨胀。监管机构不得不制定战略，以尽快赶上企业的应用发展。

但监管机构也开始关注技术偏见。在其2016年报告《大数据是包容性工具还是排他性工具？》（*Big Data: A Tool for Inclusion or Exclusion?*）中（bit.ly/2WAvkf2），美国联邦贸易委员会（FTC）表示：“我们正处于大数据时代。

现在几乎人手一部智能手机、每个家庭都配有一台电脑，连接到互联网的设备数量不断增加，整个经济体系内的消费者数据量也在持续快速增长。”是的，大数据正越来越大。

该报告还指出：“对这些数据的分析通常有益于公司和消费者，因为其可以指导新产品和服务的开发，预测个人偏好，帮助开展定制化服务并指导个性化营销。与此同时，部分质疑人士、学者等也担心大数据分析的某些用法是否会损害消费者利益，尤其是低收入和服务匮乏人群。”

大数据的危害

就这一点而言，大数据应用必然藏着风险。FTC报告认为大数据的使用涵盖了“广泛的分析”，但2016年的摘要更侧重于消费者数据的商业应用及其对低收入和服务匮乏人群的影响。这些人并不是唯一可能因数据、软件或其分析中的偏见而受到伤害的人群。[欲了解更多FTC行动的信息，可参考“算法问责”（Algorithmic Accountability）]

FTC只是众多研究商业活动中技术和数据带来偏见的团体之一，还有其他一些关于技术应用导致歧视和排他的研究案例。“预测警务”（Predictive policing，使用计算机算法研究先前犯罪活动数据，以确定下一次犯罪可能发生的地点和时间）正在使用新的算法处理帮派斗殴数据，这是第一个专注于帮派暴力的人工智能。有些人认为该项目录入的数据不充分，容易出现人为失误。其实根本不需要数据科学家去调查为何这种数据分析和预测的结果不理想，我们从新闻头条就能知道法院里多的是冤假错案，那么，为什么这一问题会恶化呢？

墨尔本大学开展的生物识别镜（Biometric Mirror）项目，让我想起了围绕苹果手机面部识别功能所引发的与偏见相关的争议。生物识别镜使用人工智能分析人脸及其14个特征，包括年龄、种族和吸引力程度等。为训练人工智能，墨

尔本大学研究人员要求人类志愿者对照这些特征来“判断”成千上万张照片。当人们站在生物识别镜前面时，人工智能会使用来自人类志愿者的判断数据分析其面部。鉴于人类志愿者所提供数据的主观性，我们认为其必然存在偏见，人工智能驱动的生物识别镜给出的结果也是如此。很容易就能发现，在评价员工工作或志愿者时，这种生物识别技术会被误用。

在很多场景下，人工智能可能被用于消除偏见或歧视，对此我大力支持。但社会的“阿喀琉斯之踵”也会出现：偏见会延续，所做的决策会对公众的生计、福祉甚至安全产生负面影响。幸运的是，监管机构正在不断观察和学习，但这还不够。

会计师的责任

会计专业人士必须理解人工智能及其潜在的正负两方面影响。作为数据的保存者和对良好的数据治理负有部分责任的人，会计师应该关心其数据、软件工具和分析方法是否将偏见带进了工作。也就是说，财会专业人士需要紧跟人工智能和分析技术的发展变化。

人工智能已经度过了“婴儿期”。当它遵循明确的、基于逻辑的规则以形成决策或建议时，例如人工智能在计算机国际象棋中的使用，被视作人工智能的“第一波浪潮”。人工智能的“第二波浪潮”是使用复杂的统计学习来得出解决问题的答案，如图像识别系统。人工智能的“第三次浪潮”目前还处于前沿阶段，未来其将不仅拥有第二波浪潮中的能力，还将解释决定背后的逻辑或推理。换句话说，它会告诉你它看到的图像是一架飞机以及为什么它认为是一架飞机。人工智能对会计和财务职能的潜在促进作用是无止境的。

但正如生物识别镜案例所展示的那样，在复杂的软件和人工智能算法中，总会存在误用、偏见和歧视风险。在对该领域研究的过程中，我偶然发现了IBM为降低算法中的偏见风险所做的努力，即“供应商符合性声明”（SDoC）。一份公开的SDoC报告旨在使人工智能的应用更加安全、透明和公平。这些报告展现了人工智能算法在性能标准化测试、公平性和风险因素下的表现，我不确定应该用何种标准或尺度来衡量其表现，或者是否有能够评判该领域最佳实践的权威机构。虽然IBM在技术创新领域享有很高的声誉，但或许更加中立客观的标准制定者才能让人信服。

算法问责

人工智能结合了复杂的算法和数据，但后两个部分都有可能引入偏见。监管者和监督机构已经注意到这一点，并正在努力确定如何保护公众利益免受技术偏见带来的负面影响。例如，FTC于2018年11月举行听证会，讨论了算法、人工智能和预测分析中的消费者保护问题。这包括该类技术的应用、围绕技术应用的相关道德议题、技术应用对企业和行业竞争态势的影响，以及与技术应用相关的政策、创新和市场等问题。

2019年4月，《算法问责法案》（Algorithmic Accountability Act）提交美国参议院。拟议的法案要求使用算法驱动软件的公司确保偏见不会嵌入到软件中，并在发现偏见时及时纠正。同时提出，FTC将负责监督这项工作并确保合规，其将在两年内制定相关法规，适用于年收入超过5000万美元或数据收集量超过100万人的公司。然而，该法案的反对者认为，对人工智能过度监管会抑制创新。

算法偏见

正如《算法问责法案》和FTC规范算法的尝试所示，复杂算法中的偏见也是潜在偏见的主要来源之一。这与算法使用的数据无关，相反，这与从一堆数据中分析得出结论的复杂数学公式有关。在工作中使用算法的数据科学家必须保证在管理偏见或使偏见影响最小化的同时，有信心进行预测。

数据偏见

当数据科学家用来“训练”算法的数据不能准确地反映算法模型需要解决和预测的业务问题时，就会发生数据偏见。数据将具有不同程度的复杂性和结构。具有较高方差的算法可以利用更复杂的数据，但鱼和熊掌不可兼得，如果数据超出了训练的数据集，这些算法的可信度就存疑。例如，如果自动驾驶汽车仅在白天或城市环境中接受驾驶训练，那么当在夜间或农村行驶时，它就缺少关键的



训练数据。

固化思维偏见

固化思维偏差与算法训练数据中的文化和社会问题密切相关。当在训练数据中使用关于某个群体的固化思维时，就会出现偏见。数据科学家必须确保训练数据不带有这些固化思维或偏见。举例来说，算法使用的固化思维偏见，可能是所有戴眼镜的人都更聪明或所有家庭都包括父亲和母亲。数据科学家需要了解如何避免将自己对社会的固化思维偏见引入到训练数据中。

测量偏见

测量偏见常被认为是由故障机器或设备（例如计算机或电子设备）导致的错误测量结果。这种错误的测量通常会扭曲数据输出，然后数据输出会被误解或导致错误的见解并驱动形成无效的策略。必须一开始就正确定义数据本身，例如，在一项调查中，通过询问受访者一些主要问题而生成的设计不当的数据集，将导致数据集和测量结果出现偏见，从而误导决策。

展望未来

以上这些意味着什么？通过确保技术与复杂的分析方法不会将偏见和歧视引入决策流程，管理会计师可以在保护公众利益方面发挥主导作用。这首先要理解这些技术（黑匣子里在发生什么）以及偏见可能在哪里出现（包括算法本身）；要审查数据生命周期中的人际互动，从而将偏见最小化甚至消除；需要制定完善的数据管理战略和政策。同时，必须开发无偏见的技术和算法。会计师们也需要花时间掌握所需的数据科学知识及分析能力，这样才能在同行中脱颖而出，为组织创造更多价值。会计师必须具备技术和人工智能方面的技能——这不是“锦上添花”的技能，必须知道如何消除黑匣子里的偏见风险，进而有效管理数据，并朝着无偏见的洞察、战略和决策方向发展。

FTC报告推荐了几种途径来预防偏见，避免美国法律下的法律和道德风险：

- 1.检查数据集以及分析中使用的基础算法，以确保隐藏的偏见不会对某些利益相关方群体产生意外影响。
- 2.请牢记：对大数据进行分析可能会发现相关性，但这并不意味着相关性对业务有意义。你应该平衡使用这些结果的风险，特别是在你的政策可能对某些利益相关方产生负面影响的情况下。当大数据工具用于重要决策时，例如那些影响健康、信贷和就业形势的决策，有必要对数据和算法进行人为监督。
- 3.应考虑在某些特定情境下，是否要为维护公平和道德底线而放弃使用大数据。
- 4.应确定你是否可以使用大数据来为先前未被充分代表的利益相关者提供机会。

这对管理会计师来说似乎是个不错的起点！**SF**

Brad J. Monterio, IMA全球董事会成员, IMA技术解决方案和实践委员会名誉主席及高级顾问、多元化和包容性委员会主席, Colcomgroup公司（一家位于纽约的咨询公司）常务董事。联系方式：bmonterio@colcomgroup.com。

“我与 CMA” 征稿

十年可能很长，长到让你达到事业的巅峰；
十年可能很短，短到你在职场起起伏伏，
依然初心未改。IMA 愿意做你的倾听者，
与你一起回忆走过的路，低头看脚下的路，
展望未来的路。



转眼间，CMA® 美国注册管理会计师中文考试已经推出 10 年了。您可能早早就把 CMA 证书握在手中，从刚入职场的新人成长为业务熟练的老生；您也可能初识 CMA，正在思考要不要报名考试，埋头苦读。不管怎样，我们想看看您眼中的 CMA 是怎样的，听听您与 CMA 考试的相爱相杀，并与您一同展望中国管理会计的未来。

IMA® 美国管理会计师协会现组织征稿，主题为“我与 CMA”，可围绕以下主题书写您的经历、感悟和建议：

1. 从听到 CMA 到考取 CMA 证书的心路历程；
2. CMA 证书对您工作、学习、生活的影响；
3. CMA 的知识体系对您的工作有哪些具体帮助；
4. 您供职企业运用管理会计（或 CMA 知识体系）

运用的案例；

5. 您所在单位、学校等对 CMA 证书是否认可与支持；
 6. IMA 如何推动中国管理会计的发展；
 7. 对 IMA 的建议，如考试安排、继续教育、出版物等；
- 稿件内容不局限于以上主题。稿件字数不超过 3000 字。如稿件被采用，作者将获得每篇 600 元的稿酬。

本征稿长期有效，随时欢迎投稿。请将稿件发送至 yfeng@imanet.org。稿件被采用后，将于 IMA 的内部刊物《战略财务》上刊登，或者在“CMA 官微”公众号上登载。