

# 基于大数据的审计技术创新思路

张书慧<sup>1</sup>(副教授), 胡士雄<sup>2</sup>

**【摘要】** 大数据浪潮汹涌而来,而数据正是审计赖以生存的基础。文章从分析大数据和审计技术的交集入手,介绍了大数据条件下的审计技术创新路径,并对大数据在审计领域中的应用困境进行了归纳总结。

**【关键词】** 大数据; 审计技术; 创新路径

**【中图分类号】** F239.1

**【文献标识码】** A

**【文章编号】** 1004-0994(2016)04-0066-2

## 一、大数据和审计技术的交集

认识“大数据”至少需要弄清楚两个问题:一是数据和信息的区别,二是大数据怎么才算大。首先来看第一个问题,多数情况下,“数据”和“信息”两个词可被替换使用,但从严格意义上讲,它们是有很大区别的。数据是对信息数字化的记录,其本身并没有意义,而信息是指把数据放置到一定的背景下,对数字进行解释,赋予意义。例如:“16”是个数据,而“辽宁舰的舷号是16”则是一条信息。进入信息时代后,人们趋向把存储在计算机上的信息,无论是数字、音乐还是视频,都统称为数据。再来看第二个问题,大数据的大肯定是超出了传统意义上的度量单位的度量范围,但到底多大才算得上大呢?一般认为,大数据的数量级应该是“太字节”(2<sup>40</sup>),而麦肯锡全球研究所认为,并不需要给“什么是大”定出一个具体的“尺寸”,因为随着技术进步,这个“尺寸”还在不断增大。

大数据需要大量的数据,而经济领域会产生大量数据,虽然在微观经济领域,记录个体行为的数据似乎是杂乱无章的,但当数据累积到一定程度时,不论是个体还是群体的行为都会在数据上呈现一种秩序和规律。审计是干什么的?审计是由独立的专门机构和人员,依法对被审计者的财务、经济数据和其他资料所反映的经济活动的真实性、合法性和效益性进行审查,并得出结论的一种监督、鉴证和评价活动。从这个定义上我们就能看出,审计工作是对数据进行挖掘和分析,进而找出问题。日常审计中,我们一直在使用和研究的审计抽样技术,一方面是由于对审计成本的考虑,另一方面也反映了对大数据的无奈。然而,随着信息技术的进步,大数据的应用成为了可能,它无疑给我们未来的审计技术市场开拓了更加广阔的前景。

以社会公共福利为例,审计发现的问题很多,如虚报账单、隐瞒收入和存款,更可恶的是装穷吃低保、分福利房和保障房。如果仅仅靠公示制度,由于信息不对称,普通公民并不具备监督的条件,而站在国家审计层面,我们可以通过对审计对象的收入、家庭条件及其日常活动轨迹等数据的收集,通过计算机自动确定相互矛盾、异于常态的记录,一旦发现造假或者不实的数据,则转入人工追讨环节,这样不仅可以节约时间,而且也将使违规资金无处遁形。

## 二、大数据条件下的审计技术创新

审计离不开数据,而互联网使数据呈现爆炸式的增长。总体来看,我们接触的大数据有四个来源:过程数据,如会计经济业务留下的运行轨迹;环境数据,如大气参数、人体指标等;社会行为数据,如人们使用微信、乘坐交通工具等产生的记录;物理实体数据,如房产、机器和材料的特性数据。有了数据我们才能展开审计技术创新路径研究。

**1. 数据联接——分析多维。**联机分析本意是把分立的数据库“相关联”,进行多维分析。“维”是联机分析的核心概念,指的是人们观察事物、计算数据的特定角度。比如中国足球超级联赛,如果要分析某支足球队联赛状况,可以按进球多少、控球时间、全场跑动里程进行分析,这些不同的分析角度就叫“维度”。分析问题的角度,可以视为一个或多个维度的交叉。举一个例子:营业税的缴纳涉及两个方面,一方面是开票的收入,另一方面是营业税的金额。如果把这两个数据库联机分析,营业税总额在收入总和中应占一定的比重,即营业税税率要在一定的区间内才是合理的,否则为异常。如果再进一步细分,按照公司注册号来进行匹配,对微观主体进行分析也能发现缴纳营业税是否存在问题。不论是从时间、空间角度,还是从其他角度,都能找到一个合乎逻辑的解释。有

**【基金项目】** 河北省审计厅2015年重点科研课题“数据化条件下的审计技术方法创新研究”(课题号:201511)

了联机分析,审计人员就可以根据自己的需求,从不同的维度、不同的粒度进行数据分析,从而获得全面、动态、可随时加总或细分的分析结果。

**2. 数据挖掘——豁然开朗。**如果说联机分析是对数据的一种透视性的探测,数据挖掘则是对数据进行挖山凿矿式的开采,它的主要目的有两个:一是发现数据表面下的潜在规律;二是通过数据回归,对未来进行预测。前者称为描述性分析,后者称为预测性分析。数据挖掘一直是传统审计中经常应用的技术,在结构化数据库中,通过筛选和排序,找到潜在的关系和规律。但随着舞弊技术的不断升级,越来越多的虚假信息散见于各种数据之中,隐蔽性更强,只有万不得已,我们才会采取人海战术,把成千上万的数据进行一一核对,代价无疑是惊人的。因此,我们要创新审计技术,高端的数据挖掘开始出现端倪。例如,通过网络留言挖掘网民的意见。网民在博客、论坛、社交网站留言发表的看法和评价,是一种非结构化的数据,把散布在网络上的这些资源整合,从中挖掘到有价值的信息和线索,正是我们应尝试去做的事情。群众的力量和智慧是无穷的,在审计中,我们关心的问题,很多也是他们关心的,其中不乏相关领域的专家、能手,通过网络平台听取他们的意见,审计工作或许会豁然开朗。

**3. 数据追踪——有力震慑。**数据挖掘是对存在的数据进行分析,从而揭示数据之间隐藏的关系、模式和趋势,为决策者提供新的依据。而数据追踪则是对数据产生的源头进行实时跟踪,留下它应该留下的不同时间点的图像、声音,甚至是视频记录,这些数据为得出结论提供了有力的支撑。比如农村电网改造审计,其中审计的重点无非两个:一是工程物资的真实性和入账问题;二是占地赔偿费问题。后者在实际中经常会引发争议,还容易出现截留挪用。作为事后审计,往往无法还原当时的场景和状态,比如砍了几棵树、毁了多少苗,项目完工进行结算审计时,当时留下的记录都无法再还原。如果能够进行数据追踪,留下征地补偿时现场作业的图形影像,甚至是谈话记录,可能审计起来就不会产生争议,也不用过多解释,结论会让人心服口服,这样也可极大地遏制农村电力工程项目上的财务舞弊行为,给企图侵占国家财产的人以震慑。再比如,对于敏感的领域,如资金结算处、重要会议等,是不是应该加装监控,这样方便别人的同时也方便了自己。数据追踪,要的是源头的真实准确,如果源头找到了、过程明晰了,结论当然也就水到渠成了。此时,审计不是变得简单了,而是变得更有监督力度了。

**4. 数据归类——规律可循。**审计时,按照一定的数据属性归类,特别是在项目审计中,就会有意想不到的效果。比如我们在全中国大数据中发现同是一个地区,建造两座大小相同的水库,成本差异一般较小,或者通过大数据发现相似的地区建设水库的成本差异也应该也较小,而实际被审计的建设水库与归类的其他水库的数据差异较大,又不能合理解释,

这就要引起审计人员的注意了。再比如,通过大数据我们以身份证为信息字段,关联银行存款、车房状况、证券投资状况等,就可以很好地避免和解决不符合条件者住保障房、廉租房,甚至是吃低保、享受国家扶贫助困政策的问题。数据归类就是在大数据中找到相同或相似的情况,大数据包含的信息是海量的,应用数据归类方法,遵循事物发展变化的规律,一些问题也就在大数据的指引下找到了答案。

### 三、大数据在审计领域的应用困境

**1. 大数据方面。**现阶段,我国大数据的采集、应用和发达国家相比,还很欠缺。虽然我国的网民数量已经跃居世界第一,我们的手机拥有量已经远远超过其他国家,但是我们的信息数据量与发达国家仍相距甚远。在信息的使用上,美国自1965年起就制定了一系列关于数据信息的法案,如1965年的《信息自由法》、1974年的《隐私法》、1975年的《信息自由法修正案》、1976年的《阳光政府法》、1996年的《电子信息自由法》、2000年的《数据质量法》、2007年的《开放政府法》等等。相比之下,在法律层面我国还很欠缺。在文化方面,“差不多就行”、“难得糊涂”这些人生哲学多年来一直影响着我们的国民,大数据环境下,不求甚解是要不得的。如果说收集数据是一种意识,使用数据是一种文化、一种习惯,那么开放数据就是一种态度,互联网的出现已经使大数据应用成为一种不可逆转的潮流。

**2. 审计方面。**近年来,国家审计署要求各地建立审计数据分析平台、审计数据中心和指挥中心,将审计对象库、审计数据库、知识资源库融入该体系,形成财政资金审计全覆盖新机制。事实证明,大数据的确在审计的相关领域发挥着越来越重要的作用,然而,现阶段我国的数据规模还较小,全国性的联网数据共享系统还没有建立起来,这些都给创新审计技术的运用提出了挑战。

信息时代,无论是个人的日常消费等琐碎小事,还是事关健康、教育的重大决策,都会在各种各样的信息系统当中留下“数据脚印”。这些“数据脚印”,不论是小如芝麻,还是大如西瓜,独立保存在不同系统中可能影响不大,但是如果建立起中央数据库,通过网络进行信息整合,就可以再现一个人的生活全景,各个系统之间的数据可以相互印证,互相解释,个人隐私也无所遁形。审计对数据有着天然的依赖,这些数据脚印给普通公民以法律上的保护,对于犯罪分子则是巨大的震慑。

#### 主要参考文献:

- 徐子沛.大数据[M].南宁:广西师范大学出版社,2015.  
周洁.论大数据对会计核算原则的冲击[J].财会月刊,2015(25).

作者单位:1.河北女子职业技术学院经济管理系,石家庄050091; 2.北京鹏龙天创物资有限公司,北京100021