

可视化技术在大数据审计中的应用

——以某制造企业 IPO 审计项目为例

薛杰 杨逸云 张敏

摘要：大数据审计目前尚处于萌芽阶段，如何引进先进技术、完善大数据审计框架是亟待解决的问题。本文以某化工类制造业企业 IPO 审计项目为例，介绍可视化技术在大数据审计中的应用问题，通过展示数据采集和预处理、可视化分析、可视化展示等过程，直观呈现审计全过程，为大数据审计实践工作的开展提供参考。

关键词：可视化技术；大数据审计；可视化分析与展示

中图分类号：F239 **文献标志码：**A **文章编号：**1003-286X(2021) 21-0058-07

随着审计业务日益复杂、审计风险不断上升，传统审计模式的局限性越来越明显，向大数据审计转型是大势所趋。目前迫切需要解决的问题是如何尽快将具体的新技术应用于审计实践，逐渐搭建起大数据审计框架。可视化技术是一个简单而又实用的技术，必将大量应用于审计实践中。本文以某制造企业 IPO 审计项目为例，系统介绍可视化技术在大数据审计中的

应用问题。

一、可视化技术在大数据审计中的应用方向

（一）可视化人员进度安排

技术的进步将带来审计人员的专业化，未来审计的工作形式将从垂直结构变为平台型结构。在目前的垂直型审计模式当中，每个审计团队将独立完成业务承接、执行审计程序、出具审计报告的全过程，团队之间几乎没有联系。在这种模式当中，团队成员根据入职年限、工作经验和能力分工完成不同科目，每个人都涉及到审计沟通、取证、数据分析等所有工作，将每个成员都锻炼成了“多面手”，但就团队整体而言，这种形式不利于发挥不同成员的比较优势，无法形成专业化运营，不利于工作效率和准确性的提升。

随着大数据审计的发展，数据分析日益成为审计工作的核心，现场协调人员、质量控制人员、技术人员都将围绕数据分析工作提供帮助。因此，审计工作的形式将从现有的垂直结构转变为以数据分析平台为核心的平台型结构（张敏，2020）。可以根据人员特征将所有审

计人员分为数据分析、质量控制、现场核查、技术支持、协调沟通等多个小组，通过专业化分工提高审计工作的专业性和效率。目前一些会计师事务所已经将部分基础性工作集成到内部商业服务中心，成为专业化分工的雏形。在审计过程中可以通过仪表盘、甘特图等形式展示各平台型小组的分工情况、进度安排、遇到的问题和困难等，帮助项目负责人综合了解项目进度并协调各小组的工作，实现项目组内部的实时协调、资源共享。

（二）可视化数据建模

随着可视化技术的发展，审计工作必将日益依赖图表展示和表达结果。可视化工具能够提供直观、简洁的方法表示审计分析的过程，帮助审计人员定位可疑、重要的信息（李强和谢汶莉，2016）。

1. 词云图可以将出现频率比较高的关键词进行大小、颜色上的突出显示。在实务审计工作中，存在大量的文本数据，如销售、采购合同、企业内部文件等，审计人员可以借助词云图及时找到审计方向，定位需要进行额外审计程序的主要企业（牛艳芳等，2016）。

作者简介：薛杰，中国人民大学商学院博士研究生；
杨逸云，中国人寿保险股份有限公司；
张敏，中国人民大学商学院会计系教授，博士生导师。

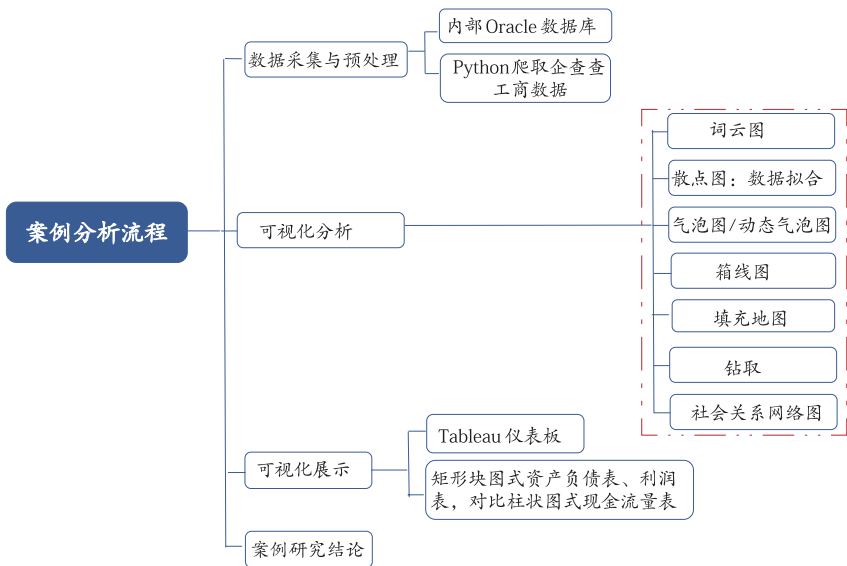


图1 案例分析流程

2. 散点图常用来表示X、Y坐标轴之间数据的变化关系,在可视化软件中可以将三个维度和两个度量相结合,表示出更多的信息。散点图还有另一个重要功能就是数据拟合。在数据拟合的过程中,通过判断横纵坐标轴大致数量关系,选择需要添加的趋势线类型,趋势线外分布较为离散的点属于异常、可疑数据,需要审计人员进行重点分析。

3. 气泡图可以通过气泡大小直观反映数值的大小。气泡图与散点图有一定的相似之处,都是通过点的位置表示坐标轴的数量关系,也可以像彩色散点图那样对不同的气泡上色,不同之处在于气泡图可以在图中额外加一个大小变量反应数量关系。特别是对于不同时间维度下的同一类数据,可以制作动态气泡图,并追踪气泡痕迹,反映同一数据在不同时间线的变化情况,帮助进行纵向时间线的分析。

4. 箱线图是一种用于显示数据分散情况的统计图表,在箱线图中用上边界、下边界、中点、上四分位数、下四分位数五个点对一个数据集进行简单总结。审计人员通过箱体的长短判断数据的波动情况,找到异常值。

5. 填充地图是进行区域分析的重要工具。在审计过程中经常会碰到有关产品销量、销售额等地区分布的信息,审计人员可以通过制作可视化地图,直观地将地理数据和财务、业务数据相结合,展示不同地区产品的销售情况。此外,为了将不同度量的数据展示在同一图表中,可以通过标记双轴的方式将销售额、利润等维度通过颜色深浅、图形大小在同一张图表当中一起展示,帮助审计人员进行区域对比分析。

6. 钻取是非常实用的可视化联机分析处理(OLAP)分析操作,可以在利用可视化图表分析业务问题时,由宏观层面向下逐级钻取至明细数据;或者先展示明细数据,再向上钻取至汇总数据。在日常审计过程中,常遇到多层分类数据,审计人员可以通过创建分层结构,逐层分析、展示这些科目的数据结构,使审计工作更加明晰、有条理。

7. 社会网络分析是研究个体、群体或社会之间的社会关系结构及其属性的规范和方法。在审计工作当中,社会网络分析有着广泛的运用空间:其一方面可以在内控分析中通过追踪企业内部单据签字流程,绘制流程关系图,发现流

程审批缺陷,评估企业内控情况;另一方面,也可以用于关联方分析领域,通过计算机编程语言(Python)从企业信息查询网站上爬取(根据万维网网页链接,获取相关万维网资源的手段)客户的股东信息后,绘制企业与股东关系网络图,直观显示企业之间持股情况,帮助审计人员找到隐蔽的关联方关系。

(三) 可视化结果展示

仪表板是显示在单一位置的多个工作表、图形的集合,能够同时比较多种数据,还可以在仪表板上添加筛选器、突出显示等功能,实现关联数据的交互分析和展示。特别是通过联机分析功能在仪表板中实现各图表之间相互联动,排除其他非相关信息的干扰,使数据使用者能够以多维的视角对相关数据进行分析并帮助其做出决策。

审计人员还可以利用可视化技术出具财务报表,增强财务报告的可读性。例如借助矩形块图实现资产负债表、利润表的可视化,用于替代资产和销售百分比法,通过矩形图块的颜色深浅、大小显示固定资产、银行存款、营业成本、销售费用、净利润等主要科目的金额和占资产或收入的比重(管彦庆等,2014),使报表使用者一目了然。可采用对比柱状图的形式实现现金流量表的可视化,左侧柱形代表流出金额,右侧柱形代表流入金额,通过柱形长度直观反映流入/流出金额的大小,并通过钻取实现三级动态图表连接。

二、基于某制造企业IPO审计项目的案例分析

本文以某化工类制造企业IPO审计为例。审计师在进行大数据审计过程中可采用的一般性流程和方法见图1。首先,从企业内部Oracle数据库中提取2017~2019年度PVC、烧碱、水泥、熟料等产品的主要客户名单,完成预处理工作。借助Python爬取客户的工商登记

信息,与企业内部客户名单配对,以便后续分析工作开展。其次,将数据导入 Tableau、Gephi 等可视化工具,绘制词云图、散点图、气泡图、箱线图、条形图、折线图、填充地图等可视化图表,寻找审计线索,结合审计人员职业判断获取审计证据。最后,运用可视化仪表板展示审计的过程和结果,帮助业务主管人员了解审计工作的进展和结果,特别是借助联机分析排除其他非相关信息的干扰,对数据进行有针对性的分析。同时通过矩形块图等可视化图表展示财务报表,增强企业年报的可读性。

本文所用数据来源于企业内部和外部公开信息。其中,企业内部信息包括企业经审计后合并财务报表、序时账及 2017 ~ 2019 年 PVC、烧碱、水泥、熟料等产品主要客户、销售数量、金额、成本、毛利率等信息。部分内部信息出于保密要求,已经进行了数据调整和马赛克处理。外部信息由于数据量大且较为庞杂,仅依靠手工难以完全获取,因此利用 Python 和 RPA 等技术手段在企业信息查询站上抓取了企业客户的工商登记信息(包括法人、股东名称、注册地址、经营范围等),以便后续研究处理。在非结构化数据方面,笔者通过浏览证监会、上交所、深交所等网站,获取了如政策文件、交易所问询函、上市公司通知公告、媒体新闻、图片等非结构化数据,了解了 IPO 审计的政策精神和关注重点,以便制定大数据审计方案和开展后续研究工作。

(一) 数据采集与预处理

审计工作经常需要搜集外部网页信息,用来与企业内部数据进行核对或者协助审计人员进行分析。例如在关联方分析中需要提取企业的工商登记数据(如存续状况、法人姓名、股东及持股情况、统一社会信用代码、地址、经营范围等)。传统做法是审计人员根据 Excel 中的企业名单,逐一在国家企业信用信息



图2 Tableau 词云图示例

公示系统、天眼查、企查查等网站中查询,再将查询结果逐一复制到 Excel 中,不仅耗费时间精力,还可能忙中出错,影响后续审计结果;在大数据审计中,可以利用 Python 编写程序,自动提取外部网站当中的数据,提高效率和准确性(陈伟和孙梦蝶, 2018)。

在本案例当中,审计人员在企业内部 Oracle 数据库中提取了 2017 ~ 2019 年度 PVC、烧碱、水泥、熟料等产品的主要客户名单,共计 231 家。通过查询这些客户的存续状态、股东背景等信息,能够协助判断收入的真实性。例如,倘若发现某家新成立企业却有着大额销量,或者多家企业拥有同一个股东,那么就需要审计人员提高警惕,对这些企业做进一步核查。除了 Python 之外,审计人员还可以通过 R 语言开发工具,采集文本文件数据、网页数据、Access、Oracle 数据库数据等多种类型的数据,提高审计工作效率。然而,目前审计大数据环境尚未搭建,缺乏数据访问和共享机制,也给大数据审计工作的开展带来了一定的不便。例如部分企业信息查询网站采取了多种反爬虫技术手段,不断更改网页源代码、同一账号短时间内搜索大量数据将被封号,或者将企业名称等提取

内容变为按汉字一个一个提取,给大数据提取造成了一定的困难。因此在推进大数据审计应用的过程中,注重数据的共享互通显得尤为必要。

(二) 可视化分析

1. 利用 Tableau 软件进行可视化分析。相比于数字,图像具有清楚直观、简单方便的特点,特别是在大数据时代,面对庞杂的审计数据,审计人员更加需要通过图像来发掘、展示审计证据,提高审计工作的效率和可理解性。本案例主要选取 Tableau 软件进行可视化分析。在进行可视化分析之前,需要将经过提取和预处理后的 Excel 客户销售情况统计导入 Tableau 当中。Tableau 具有数据的实时更新功能,如果后续需要更改或添加数据,可以直接在原 Excel 文件中修改,在 Tableau 数据库中点击刷新即可。

(1) 词云图。为核实销售收入的真实性,审计人员需要从 231 家主要客户中选取销售额大、毛利高的客户作为重要客户,进行电话访谈或实地走访。在信息化环境下可通过 Excel 或结构化查询语言 (SQL) 排序筛选客户。如图 2 所示,颜色偏红代表该公司销售金额高,深蓝色代表销售金额较低,字体越大则代表毛利越高。可以看出,销售额高、



图3 Tableau 散点图示例1

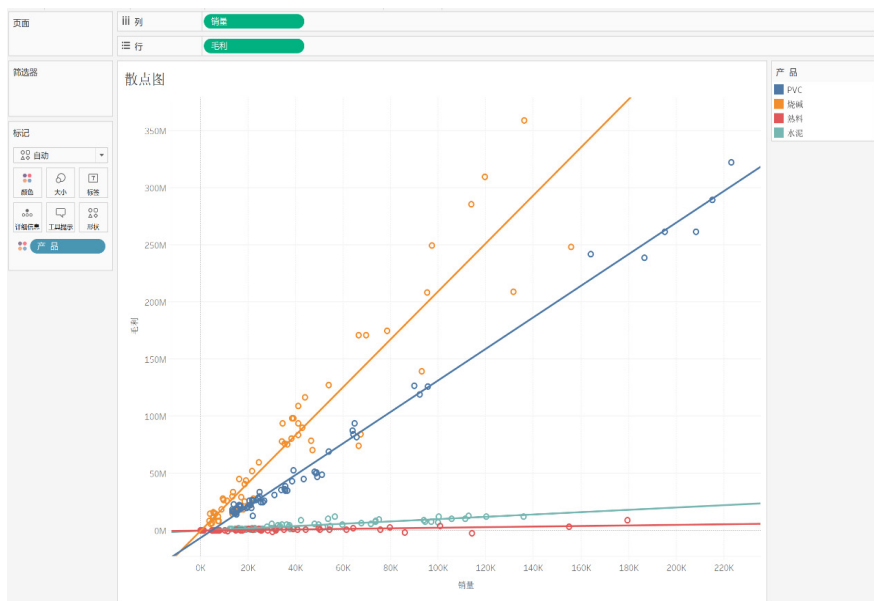


图4 Tableau 散点图示例2

毛利大的是浙江**控股集团股份有限公司、浙江**氯碱化工有限公司、山东**能源****铝业有限责任公司等。

(2) 散点图。审计人员需要按时间和产品类别，观察毛利率的变化情况。如图3所示，将“时间”和“毛利率”字段分别拖至“列”和“行”，选择散点图，便会出现以时间和毛利率为度量的横纵坐标轴。通过标记区的各项功能建立多维度的图表：将“收入”字段拖至颜色区域，

蓝色越深表示客户销售额越高；将“客户”字段拖至标签区域后可以在图中显示客户名称，便于审计人员观察图表；将“产品”字段拖拽至筛选器区域，选择PVC、烧碱、水泥、熟料等产品类型中的熟料，可以在图中筛选出熟料客户的毛利率分布数据。该多维度图表可以协助企业进行后续联动细致分析。观察图3可以看出，神木**水泥有限责任公司连续两年毛利率为负，表示这家企业连续

两年向该公司销售熟料的业务都是亏损的，这一现象值得审计人员重点关注，需要采取进一步的审计程序，判断企业向该公司的销售定价是否合理，是否存在定价不公平等问题。

此外，审计人员还可以利用散点图的数据拟合功能分析不同类型产品销售量和毛利之间的关系，将“销量”和“毛利”字段分别拖拽至“列”和“行”区域，将“产品”字段拖拽至标记区的颜色区域，得到散点图，不同颜色代表不同类别的产品品类。分析产品销量与毛利之间的线性关系，选择线性拟合。如图4所示，四种产品当中，烧碱的单位产品毛利最高，熟料的单位产品毛利最低，且除烧碱外，其他三类产品的单位产品毛利比较稳定，不同企业之间单位产品毛利值相同，而烧碱产品的单位产品毛利变化较大，特别是孝义市**化工有限公司销量高、毛利也高，属于优质客户，审计人员应当对此类企业予以重点关注。

(3) 气泡图。在IPO审计过程中，常常需要对企业3~5年的经营、财务数据进行对比，既可以采取折线图、条形图等方式呈现，也可以采取动态气泡图的方式呈现。和折线图、条形图相比，动态气泡图的优势在于既能像折线图一样准确反映数据的变化过程，又可以像条形图一样准确比较金额的大小，兼具两者之长。

在本案例中，审计人员需要了解不同产品类别近3年的销售金额变化情况。完成静态气泡图的制作之后，为了显示各产品金额的变化情况，将“时间”字段拖拽至页面区，此时界面右侧显示了不同的年份，点击“播放”可以看到气泡的动态变化过程。点击“显示历史记录”，可以选择标记或轨迹，将呈现历史数据，例如在图5中颜色越淡代表时间越久，可以清楚地发现随着时间的推移，PVC和水泥的销售金额逐年增加，而烧碱和熟料则小幅下降。

(4) 箱线图。审计人员需要分析被审企业产品售价是否公允,有无异常情况。从图6可知,随着时间的推移,PVC产品的平均售价逐渐增加,2017~2019年箱线图的箱体长度越来越短,说明售价的波动程度趋缓。需要特别注意的是,2017年广东**科技实业有限公司销售价格远低于同期其他企业,而同期销售量最高的浙江**控股集团股份有限公司却拥有最高的PVC销售价格,与销量越高、价格越低的认知相违背,可能存在合同规定其他优惠条款、销售价格不公等情况,需要审计人员对两家公司实施进一步的审计程序获取审计证据。

(5) 填充地图。在Tableau中,可以自动生成地图等区域图表。Tableau自带Mapbox、WMS服务器等地图源,可下沉至中国地图的省、自治区、直辖市以及主要城市,倘若需要下沉至区县一级,可以下载高德等地图包,手动添加至Tableau当中完成地图制作。在Tableau中创建地图的第一步是将“维度”当中的“地区”字段转变为地理角色,单击右键可选择转变为省、市等地理角色,之后文本字段会变为地球图表,同时在“度量”区域也会出现相应的经度和纬度,从而能够更加准确地实现地理定位。将经度拖动至“列”并将纬度拖至“行”区域,图表区会自动呈现地图,此时选择需要在地图中展示的字段,如将销售收入拖拽至标记区的颜色区域,在地图中便会根据销售收入的多少出现深浅不同的颜色。在审计实务中,经常需要把销售额和利润放在同一张图表当中一起展示以发现毛利率异常的区域。在地图中也可以将销售额和利润结合展示,例如可以用颜色深浅代表销售额的多少,用图像的大小来表示利润的多少。

(6) 钻取。审计人员需要对四类产品在不同区域、不同地区的销售情况进行分析。为使各省份名称和华北、东北等区域名称形成关联关系,需要先在

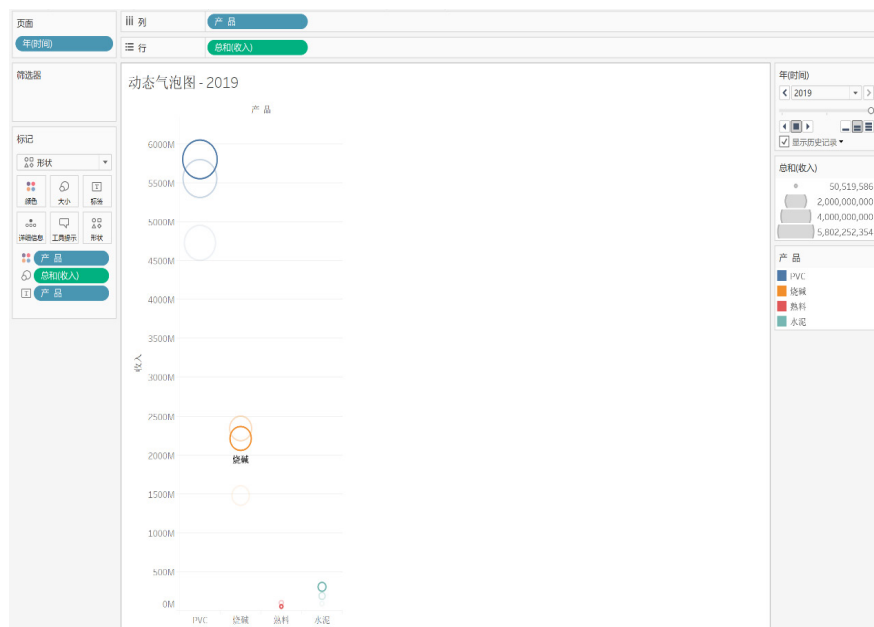


图5 Tableau 动态气泡图示例

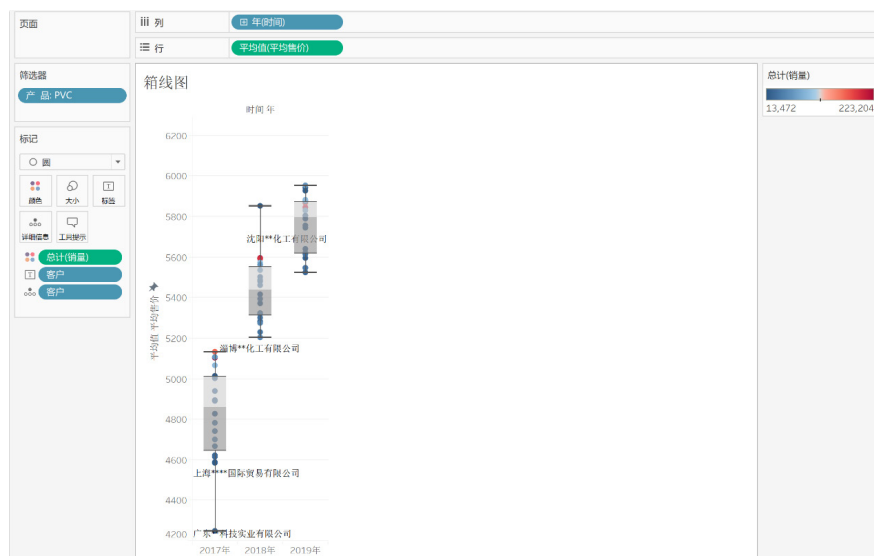


图6 Tableau 箱线图示例

Tableau中创建分层结构,具体有两种方式:一是直接将“区域”维度拖拽到“地区”维度的上方,自动弹出分层结构窗口,将新的层级命名即可(倘若有多重层级,可进行重复操作);二是选中需要分层的所有维度,选择创建分组结构。如图7所示,构建层级结构后,先将宏观层面维度“区域”拖拽至“行”区域,点击维度前方加号即可向下钻取。钻取

后,同样点击明细层面维度“地区”前方减号即可进行上卷操作。

2. 利用Gephi进行关联方分析。Gephi是一款开源的跨平台复杂网络分析软件,可以用于探索性数据分析、链式分析、社会网络分析、可视化分析等诸多方面。Gephi有概览、数据资料和预览三个界面:概览用于图形处理和可视化编辑;数据资料分为节点数据

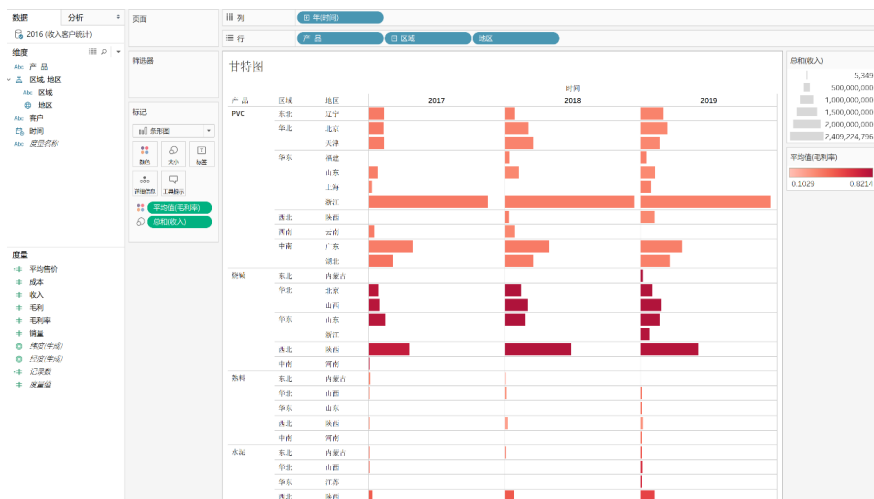


图7 Tableau 钻取操作示例

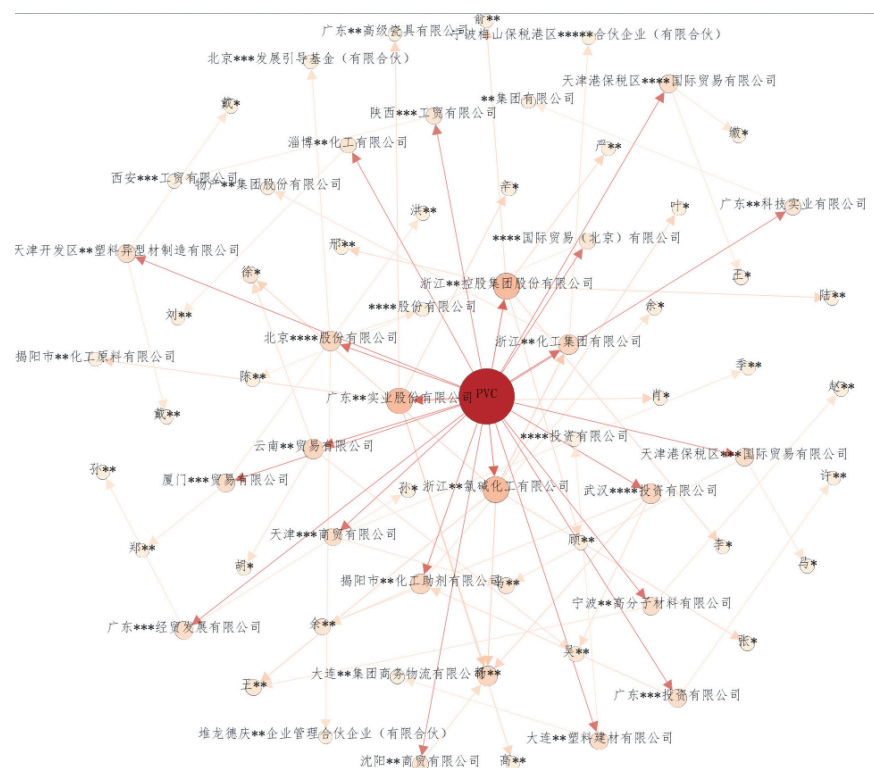


图8 Gephi 企业股东关系图示例

和连线数据两种,代表图像中的点和线;预览界面主要用于图像处理完毕后预览输出。在本案例中,审计人员结合 Python 抓取的客户、供应商股东列表,绘制 PVC 产品客户、供应商股东关

系图,找到 PVC 产品客户、供应商之间股东关系异常的企业。首先将客户和股东关系表导入 Gephi 当中, Gephi 的数据源中必须包含 Source 和 Target 数据,即社会网络当中的节点,两节点之间默

认从 Source 指向 Target。导入 Gephi 后需要对图像进行可视化编辑,通过定义节点大小、颜色、排序调整节点的显示结果,改变布局让节点分布更为均匀,随后在预览中输出图像结果。图 8 为在 Gephi 中绘制的被审企业 2019 年 PVC 客户股东关系情况。审计人员通过观察该图,发现广东***投资有限公司的股东是许**和揭阳市**化工助剂有限公司,而揭阳市**化工助剂有限公司也是被审企业 2019 年 PVC 产品的销售客户。因此,在审计过程中,审计人员需要对被审企业与揭阳市**化工助剂有限公司、广东***投资有限公司之间的业务往来重点关注,判断是否存在异常业务往来。

3. 可视化展示。在大数据审计时代,审计人员可以使用仪表板(见图 9)和可视化财务报表清楚地展示审计过程和结果。本案例使用矩形块图完成资产负债表(利润表略)的可视化。首先要对经审计后的财务报表进行数据处理,将所有资产、负债和所有者权益科目放在同一列,并增设“类别”列区分科目类别。预处理后的资产负债表导入 Tableau 后,将科目金额拖拽至标记区的颜色和大小区域,在标签区域拖入科目和金额占比。此时选择自动生成矩形块图(如图 10 所示),颜色较深、矩形块较大的为占比较大的科目。可以直观看出,该企业固定资产、应收票据、应付账款等属于占比较大的科目。

为了将所有资产类、负债类、所有者权益类科目合并,将“类别”维度拖拽至标记区详细信息区域,可以看到图片按照不同类别分为三部分,其中资产部分的面积等于负债与所有者权益部分面积之和,符合会计恒等式。此外,还可以使用钻取功能,在 Tableau 当中实现会计要素到科目甚至到明细科目、报表附注的动态链接。现金流量表可视化可以采用对比柱状图的形式,左侧柱形代



图9 Tableau 仪表板示例

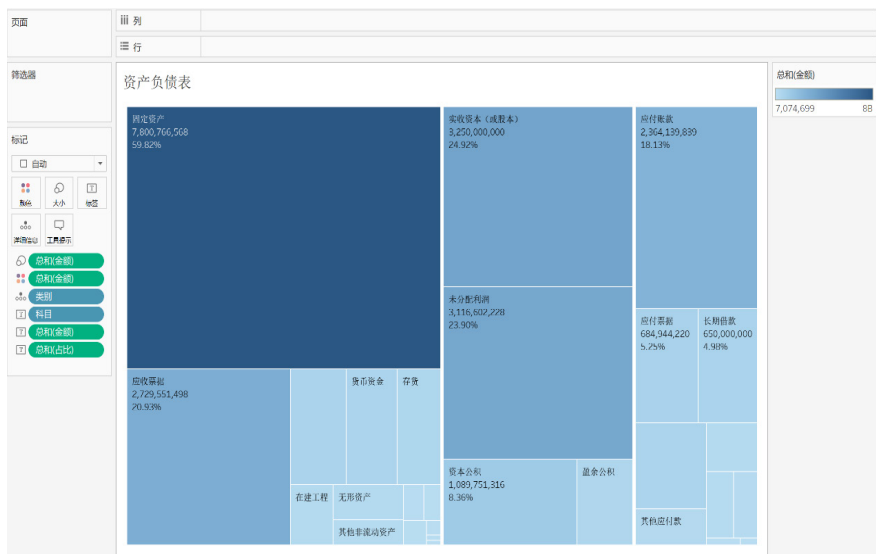


图10 Tableau 资产负债表示例

表流出金额, 右侧柱形代表流入金额, 通过柱形长度直观反映流入/流出金额的大小。

三、目前可视化技术审计的不足

审计人员可以借助 Python、RPA 等技术搜集外部结构化、非结构化数据, 增强审计证据的充分性。同时在数据分析阶段, 借助 Tableau、Gephi 等可视化

技术发现疑点数据, 挖掘审计证据, 从而提高审计效率, 更为客观地发表审计评价。近年来, 大数据技术在审计工作中的应用日益广泛, 智能审计的时代已经到来, 但目前大数据可视化技术尚未完全成熟, 在实际应用中仍存在一些问题和需要改进之处。例如本案例中涉及到数据的互联互通问题, 企查查、天眼查等信息网站采取了反爬虫设计, 增加

了审计人员利用大数据技术爬取信息的难度, 阻碍了数据的提取分析。审计人员需要不断地对原有 Python 代码进行修改更新, 才能提取所需企业的工商信息。另外对于数据量过大、可视化图表分析中信息过密、显示不够直观的问题, 需要审计人员对数据进行筛选处理, 才能保证图像数据清晰明了。此外, 在大数据审计过程中, 可视化技术更多起到辅助分析、展示的作用, 需要审计人员根据工作需要和职业判断, 选择恰当的可视化图表, 发现疑点数据, 结合其他审计程序获取审计证据。不同的可视化图表对于审计疑点的发现、效率的提高作用也不同, 未来可以在搭建集成性大数据可视化审计平台时对常用可视化图表设置提示和推荐功能, 方便审计人员选择。

责任编辑 李斐然

主要参考文献

- [1] 鲍朔望. 大数据环境下政府采购审计思路和技术方法探讨[J]. 审计研究, 2016, (6): 13-18.
- [2] 陈伟, 孙梦蝶. 基于网络爬虫技术的大数据审计方法研究[J]. 中国注册会计师, 2018, (7): 76-80.
- [3] 管彦庆, 杨喜梅, 博斌. 我国企业财务报告的可视化研究——基于上市公司合并资产负债表的 Treemaps 图形化视角[J]. 中国注册会计师, 2014, (9): 74-79.
- [4] 李强, 谢汶莉. 大数据审计中的可视分析[J]. 中国内部审计, 2016, (2): 79-86.
- [5] 牛艳芳, 邓雪梅, 陈伟. 数据科学工具之 R 语言在审计数据分析中的应用探索[J]. 中国注册会计师, 2016, (9): 93-97.