

基于大数据可视化分析技术的大气污染防治审计方法研究

陈伟 高嘉文

摘要：污染防治审计是目前审计领域关注的重点，大气污染防治审计是污染防治审计的一项重要内容，大数据技术的发展为解决这一问题提供了机遇。本文首先分析了研究大气污染防治审计方法的重要性、目前常用审计方法及其存在的不足，在分析了常用大数据审计技术方法的基础上，提出了基于大数据可视化技术的大气污染防治审计方法，并分析了该方法的原理。最后以某大气污染防治审计项目为例，验证了该方法的有效性。

关键词：大数据审计；大气污染防治审计；数据可视化技术；数据分析

中图分类号：F239.1, C931.6 **文献标志码：**A **文章编号：**1003-286X (2019) 07-0065-04

一、引言

研究如何应用大数据技术开展污染防治审计具有重要的社会价值和应用价值。大气污染防治、水污染防治、土壤污染防治等都是污染防治攻坚战的重要内

容。笔者以大气污染防治为例，结合目前大数据技术以及大数据审计的研究与应用现状，研究基于大数据可视化技术的大气污染防治审计方法。

二、大数据审计技术方法分析

目前，大数据审计技术方法的研究与应用一般集中在以下三个方面：

（一）大数据智能分析技术

大数据智能分析技术以各种高性能处理算法、智能搜索与挖掘算法等为主要研究内容。这是目前大数据分析领域的研究主流，它从计算机的视角出发，强调计算机的计算能力和人工智能，例如各类面向大数据的机器学习和数据挖掘方法等。目前关于大数据智能分析技术的研究在审计领域的应用仍不成熟，多是停留在理论研究层面。

（二）大数据多数据源综合分析技术

大数据多数据源综合分析技术是通过采集来的各行、各业、各类大数据，采用数据查询等常用方法或其他大数据技术方法进行相关数据的综合比对和关联分析，从而发现更多隐藏的审计线索。大数据多数据源综合分析技术是目前审计领域应用大数据比较成熟和主流

的内容。

（三）大数据可视化分析技术

大数据可视化分析技术是目前大数据审计应用比较成熟和主流的技术。常见的大数据可视化技术包括：时间序列图、柱状图、折线图、饼图、散点图、气泡图、雷达图、地区分布图、树地图、热力图，以及文本可视化技术标签云等。本文根据大气污染防治审计的需要，选择了适合本审计项目需要的时间序列图、散点图、地区分布图、标签云等可视化技术。

三、基于大数据可视化技术的大气污染防治审计方法原理分析

（一）大数据环境下大气污染防治审计所需的数据

大数据环境为大气污染防治审计提供了全方位分析的相关数据，比如，审计人员可以对以下数据进行分析：

1. 文本数据。这类文本数据如被审计单位的业务介绍、部门年度工作总结、批准的相关项目、大气污染防治工作情况的报告、相关审计报告。通过这些文本数据，审计人员可以了解目前被审计单位的相关业务情况、大气污染防治工

基金项目：国家自然科学基金(71572080)

作者简介：陈伟，南京审计大学教授，博士，江苏省审计信息工程重点实验室副主任；高嘉文，南京审计大学硕士研究生。

作情况等,便于审计人员开展相关审计工作。

2. 空气质量监测数据。这类数据如空气质量日均值监测数据、国控站点和省控站点细颗粒物(PM2.5)小时监测数据等。通过这类数据,审计人员可以掌握目前被审计地区空气质量等相关信息。

3. 站点经纬度信息。这类数据如国控和省控站点的经纬度信息,审计人员可以使用这些数据来查看国控站点和省控站点的分布情况,确定国控和省控站点的位置,从而为相邻站点监测数据分析比较等打下基础。

4. 相关企业用电数据。这类数据用于分析相关企业的生产情况,审计人员可以使用这些数据来分析相关企业与大气污染物排放之间的关系。

5. 相关财务数据。这类数据如大气污染防治资金使用情况等,审计人员可以使用这些数据来分析被审计单位是否合理使用了大气污染防治资金。

6. 其他外部相关数据。

除了通过以上方法获得被审计单位的相关数据之外,审计人员还可以通过一些大数据采集工具(如网络爬虫等)抓取相关环保部门网站上的相关公开空气质量监测数据,便于审计人员辅助判断空气质量监测数据的真实性等情况以及开展其他大数据分析。

(二) 基于大数据可视化分析技术的大气污染防治审计方法原理

基于大数据可视化分析技术的大气污染防治审计方法原理可概述为:根据对被审计单位的调查,在访谈和现场观察等基础上,采集被审计单位的内外部相关大数据如相关财务数据、站点经纬度数据、相关企业用电数据、空气质量监测数据等结构化数据,被审计单位批准的相关项目数据、大气污染防治工作情况的报告、部门年度工作总结等相关数据等非结构化数据,

然后对采集来的相关数据进行审计大数据预处理,在此基础上,基于“总体分析、发现疑点、分散核查、系统研究”的审计思路,采用大数据工具对空气质量监测数据、相关财务数据、相关企业用电数据、被审计单位批准的相关项目数据等相关结构化和非结构化数据进行建模和整体分析,审计人员通过对可视化的分析结果进行观察,快速从被审计大数据信息中发现异常情况(如大气污染防治相关数据变化情况,被审计地区大气污染防治相关信息系统建设使用情况等),获得审计线索。在此基础上,通过对这些异常数据做进一步的延伸审计和审计事实确认,最终获得审计证据。综上分析,基于大数据可视化技术的大气污染防治审计方法原理如图1所示。

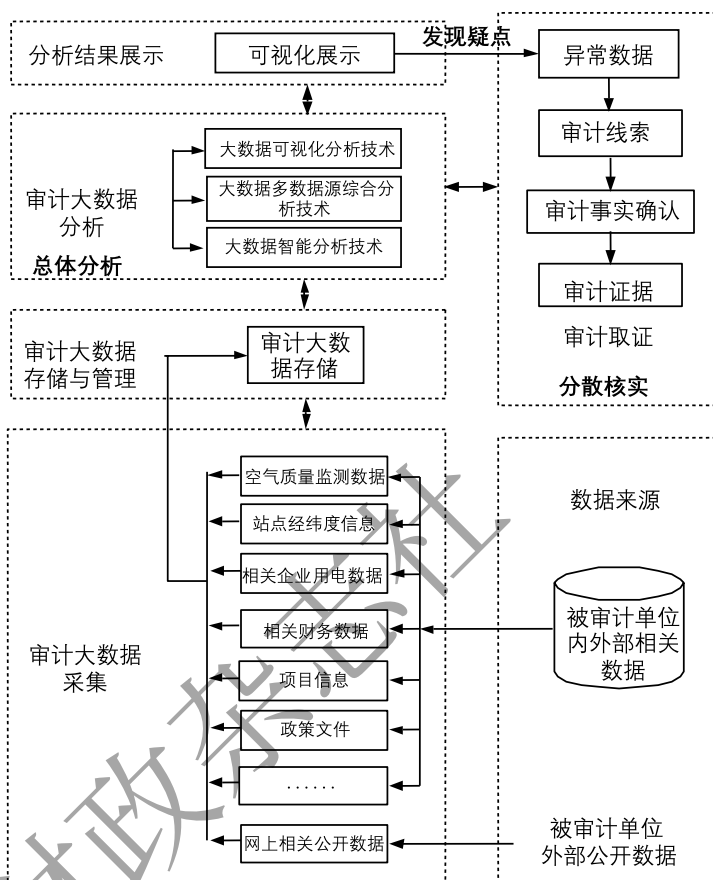


图1 基于大数据可视化技术的大气污染防治审计方法原理

四、基于大数据可视化分析技术的大气污染防治审计方法应用案例

(一) 案例背景

大气污染防治、水污染防治、土壤污染防治等都是污染防治攻坚战的重要内容。本案例以加强大气污染防治,改善大气环境质量为背景,以A地2013~2018年间大气污染防治情况审计为例,研究大数据可视化技术在该类型审计中的应用。假设现已获得A地2015~2017年空气质量日均值监测数据和A地国控站点和省控站点2017年PM2.5小时监测数据等相关大数据。

目前,大数据分析工具主要包括商业化软件工具和开源的、可编程的工具,本文以简单、流行、开源的大数据可视化分析软件R语言为例,实现大数据可

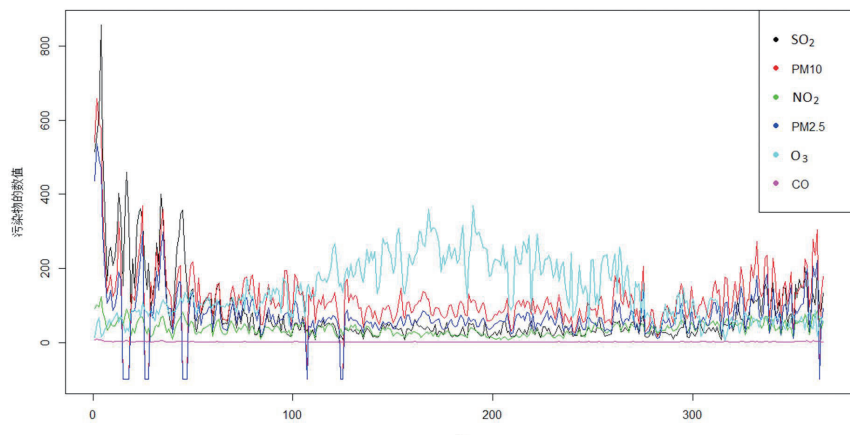


图2 六项污染物时间序列对比图

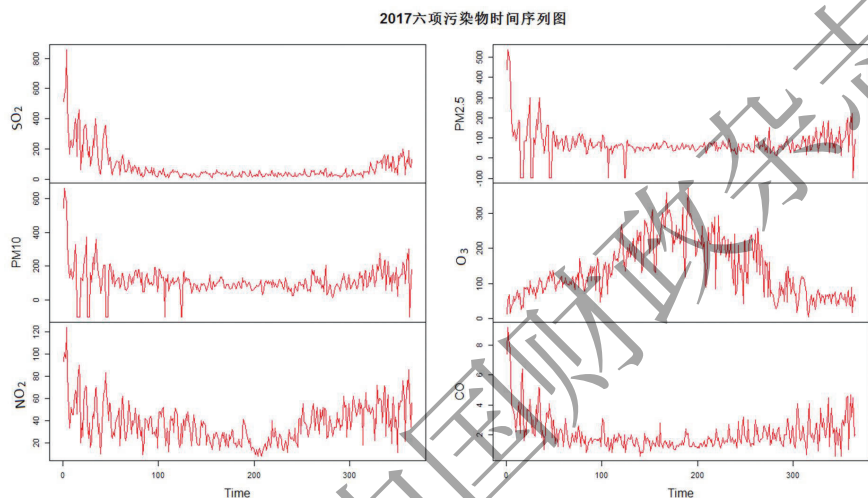


图3 六项污染物分布各自的时间序列图分析示例

可视化分析。

(二) 基于地区分布图分析的相关空气质量数据总体分析

热力图是一种可将变量值用不同的颜色或高亮形式描绘出来的数据可视化技术，它可以非常直观地呈现一些不易理解或表达的数据，如频率、密度、温度等。

为了从整体上掌握被审计地区空气质量情况，审计人员有时需要获得一些相关环保部门网站上的相关公开空气质量监测数据。在实际的审计工作中，当需要查询大量的空气质量监测数据时，

需要花费审计人员很多的时间和精力，如何批量查询相关数据成为困扰审计人员的一件重要的事情。网络爬虫工具的出现为解决审计人员的这一问题提供了一种可行的方法，审计人员可以通过网络爬虫工具采集该地区相关空气质量监测数据。在此基础上，借助R工具，采用热力图进行分析，从而非常直观地从整体上掌握被审计地区的空气质量情况。

(三) 空气质量变化情况分析

时间序列图是以时间轴为横轴，变量为纵轴，用于观察变量随时间变化情况的图，也叫推移图。在大气污染防治

审计过程中，审计人员可采用时间序列图有效地分析空气质量变化情况。比如，通过时间序列图分析，可以比较不同时期区域内环境空气质量变化情况，检测该地空气质量有没有变差；可以判断该地大气污染防治有关约束性指标的完成情况，如二氧化硫(SO₂)等排放物控制指标、PM2.5或可吸入颗粒物(PM10)年均浓度下降考核指标完成情况等；通过分析六项污染物的浓度变化情况，可以检测该地有没有篡改、伪造监测数据，严格履行监测职责。以该地2017年空气质量监测数据分析为例，为了整体把握该地2017年的空气质量变化情况，可以采用时间序列图分析2017年六项污染物数值变化情况，其分析结果如图2所示。为了更清晰地观察看每一项污染物数值变化情况，各项污染物的年度时间序列图如图3所示。其中，六项污染物分别为：二氧化硫(SO₂)、二氧化氮(NO₂)、一氧化碳(CO)、臭氧(O₃)、PM2.5和PM10。

由图2可以容易地查看2017年六项污染物在每个时间点的数值大小和变化情况。由图2和图3不难发现：除了污染物臭氧(O₃)在夏季出现数值大幅上升的情况，其余五项污染物的数值均在一月份和二月份出现大幅上升情况。其中，在一月份，二氧化硫(SO₂)的数值甚至飙到了800以上，颗粒物的数值也出现了600以上的情况，空气质量比较恶劣。另外，通过分析结果发现：PM2.5或PM10监测部分数据出现负值，大气污染物排放自动监测设备传输数据异常，需要引起审计人员的注意，应当及时进行调查，从而可能发现影响自动监测数据真实性和准确性问题的线索。

(四) 空气质量监测数据异常情况分析

在开展大气污染防治审计时，为了确保环境监测数据准确真实，防止篡改或者伪造监测数据等现象的发生，审计

人员需要对照相关法规,详细调查大气污染监测设施和系统建设情况,空气质量国控监测点位,检查空气质量自动监测点的数据是否能够真实反映辖区空气质量。为了防止部分地区通过将空气监测站设在公园,或采用棉纱堵塞采样头、向监测设备洒水等方式对空气自动监测站实施干扰,导致监测数据严重失真等问题的发生,审计人员可以通过对不同部门的或不同地区的相邻站点监测数据进行比较分析,从而发现异常监测数据线索,判断监测数据的准确真实性。比如,审计人员首先可以采用地区分布图对站点经纬度信息进行分析,查看该地区国控站点和省控站点的分布情况。在此基础上,采用散点图对比分析国控和省控相邻站点的监测数据,以某地国控和省控相邻站点监测数据分析为例,其分析结果如图4所示。

由图4不难发现:该地区国控和省控相邻站点监测数据存在不一致,需要引起审计人员的重点关注。另外,通过对该地区国控站点监测数据进行分析,发现监测数据中的异常现象(如该区的PM2.5监测数据是否明显低于同时间段内其他区域的数值),从而判断站点采样系统是否受到人为干扰。

(五) 新增大气污染物排放项目分析

为了防治大气污染,基于时间序列图等分析,对一些因大气环境质量持续恶化、二氧化硫、PM2.5或PM10等浓度长时间“爆表”的地方,需要暂停新增大气污染物排放项目的环评审批,严禁核准产能严重过剩行业新增产能项目。因此,审计人员通过对该地批准的相关项目内容数据进行分析,可以整体分析该地是否存在新增大气污染物排放项目,例如,分析该地批准的相关项目是否投向电解铝、钢铁、水泥、平板玻璃等一些高污染、高能耗等落后产能淘汰项目,如果存在这类项目,则需要审计人员对其做进一步的关注。基于以

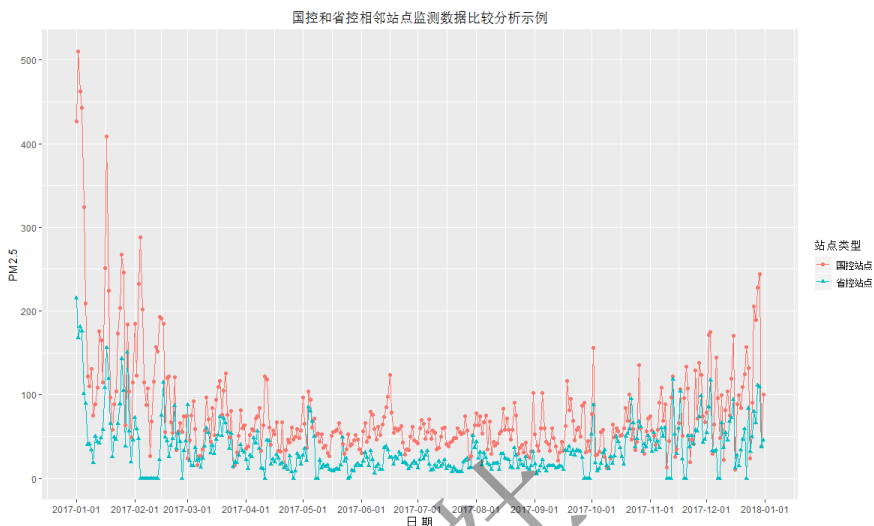


图4 国控和省控相邻站点监测数据的比较分析示例

上分析,以某地大气污染防治审计项目为例,采用标签云分析方法对该地批准的相关项目内容进行分析,其分析结果表明:该地批准的相关项目内容中存在“钢铁”“水泥”等高污染、高能耗行业的相关关键词,从而表明该地批准的相关项目中可能存在项目内容属于高污染、高能耗的行业,需要审计人员做进一步的延伸分析。

(六) 大气污染防治资金使用情况标签云分析

为了掌握大气污染防治资金的整体使用情况,确认大气污染防治资金的使用是否合理、合规和合法,需要对某一时期内大气污染防治资金使用内容进行总体分析。针对这一需要,采用标签云分析方法对从被审计单位采集来的相关大气污染防治资金支出数据进行分析,根据分析结果可以初步发现大气污染防治资金是否存在违规使用的线索。例如,采用标签云对从某地采集来的大气污染防治资金进行综合分析。不难发现,该被审计单位一部分大气污染防治资金用在了餐费、购物、烟酒等方面。根据分析出的这些线索,审计人员可以对这些大气污染防治资金数据做进一步的

详细分析,查找所有含有“餐费、购物、烟酒”等方面的支出,从而确认该被审计单位在污染防治资金使用方面是否合理、合规和合法。

责任编辑 武献杰

主要参考文献

- [1] 陈伟. 大数据技术在扶贫审计中的应用[J]. 财务与会计, 2018, (15): 69-71.
- [2] 陈伟. 审计信息化[M]. 北京: 高等教育出版社, 2017.
- [3] 陈伟. 计算机审计[M]. 北京: 中国人民大学出版社, 2017.
- [4] 陈伟, 居江宁. 基于大数据可视化技术的审计线索特征挖掘方法研究[J]. 审计研究, 2018, (1): 16-21.
- [5] 陈伟, Smiclauskas W. 大数据环境下的电子数据审计: 机遇、挑战与方法[J]. 计算机科学, 2016, (1): 8-13+34.
- [6] 陈伟, Smiclauskas W. 云计算环境下的联网审计实现方法探析[J]. 审计研究, 2012, (3): 37-44.