

税收大数据分析在工业企业疑点排查中的应用研究

王海蓉¹ 柴留祥²

1. 宁夏和宁化学有限公司 宁夏 银川 750000

2. 宁夏回族自治区自然资源宣传教育中心 宁夏 银川 750000

摘要: 本文聚焦于大数据分析在工业企业税收疑点排查中的应用。首先阐述了大数据分析的内涵与特点,以及工业企业税收疑点排查的重要性和传统方法的局限性。接着详细分析了大数据分析在工业企业税收疑点排查中的具体应用,涵盖数据采集与整合的技术手段、风险指标构建的量化方法、模型建立与应用的算法选择等。最后提出了相应的对策建议,旨在为提升工业企业税收征管质量提供有益参考。

关键词: 大数据分析; 工业企业; 税收疑点排查; 应用研究

1 引言

随着信息技术的飞速发展,大数据时代已经来临。据国际数据公司(IDC)统计,全球数据总量正以每年约40%的速度增长,预计到2025年,全球数据总量将达到175ZB。在税收征管领域,工业企业的税收征管面临着诸多挑战。工业企业通常具有生产经营规模大、业务复杂、财务核算体系庞大等特点。以某大型机械制造企业为例,其年销售收入可达数十亿元,涉及原材料采购、生产加工、产品销售等多个环节,每个环节都可能产生大量的财务和税务数据。传统的税收征管手段在面对海量数据和复杂业务时,逐渐暴露出效率低下、准确性不足等问题。如何利用大数据分析技术,提高工业企业税收疑点排查的效率和准确性,成为当前税收征管工作亟待解决的重要问题。

2 大数据分析概述

2.1 大数据的内涵与特点

大数据是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合,是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产^[1]。大数据具有“4V”特点,即Volume(大量)、Velocity(高速)、Variety(多样)、Value(价值)。大量指数据量巨大,从TB级别跃升到PB、EB乃至ZB级别;高速指数据产生和处理的速率极快;多样指数据类型繁多,包括结构化数据、半结构化数据和非结构化数据;价值指数据价值密度相对较低,但通过大数据分析技术可以挖掘出巨大的潜在价值。

2.2 大数据分析的主要技术

大数据分析涉及多种技术,主要包括数据采集与预处理技术、数据存储与管理技术、数据分析与挖掘技术

等。数据采集与预处理技术用于从各种数据源中收集数据,并对数据进行清洗、转换和集成等操作,以提高数据的质量和可用性;数据存储与管理技术用于存储和管理海量的数据,常见的存储方式包括分布式文件系统、关系型数据库和非关系型数据库等;数据分析与挖掘技术是大数据分析的核心,包括统计分析、机器学习、深度学习等方法,用于从数据中发现有价值的信息和知识。

3 工业企业税收疑点排查的重要性及传统方法局限性

3.1 工业企业税收疑点排查的重要性

工业企业是国家经济的重要支柱,其税收贡献在国家财政收入中占据重要地位。加强工业企业税收疑点排查,具有以下重要意义:一是保障国家税收收入,防止税收流失;二是维护公平公正的税收秩序,促进企业之间的公平竞争;三是提高税收征管效率,降低征管成本;四是促进工业企业依法纳税,提高企业的税收遵从度。

3.2 传统税收疑点排查方法的局限性

传统的税收疑点排查方法主要依赖于税务人员的经验判断和人工审核,存在以下局限性:一是效率低下,面对海量的工业企业数据,人工审核难以快速、全面地发现税收疑点;二是准确性不足,人工审核容易受到主观因素的影响,导致排查结果存在偏差;三是缺乏系统性,传统方法往往只能针对个别指标或问题进行排查,难以全面、系统地评估企业的税收风险;四是信息获取渠道有限,传统方法主要依赖于企业申报的数据和税务机关内部的信息,难以获取企业外部的相关信息,影响了排查的全面性和准确性。

4 大数据分析在工业企业税收疑点排查中的具体应用

4.1 数据采集与整合

4.1.1 数据来源

税务机关内部数据:包括企业申报的纳税申报表、

财务报表、发票数据、税务稽查记录等。纳税申报表数据反映了企业的纳税情况和财务状况，财务报表数据提供了企业的资产、负债、收入、费用等详细信息，发票数据记录了企业的购销业务，税务稽查记录则包含了企业过去的税收违法情况。

企业外部数据：工商登记信息可以了解企业的注册资金、经营范围、股东结构等基本信息；银行流水信息可以反映企业的资金往来情况，判断企业是否存在隐瞒收入或虚列成本的问题；电力消费信息、物流运输信息等可以反映企业的生产经营规模和投入产出情况^[2]。例如，通过分析企业的用电量，可以大致估算企业的生产产量，与企业的申报销售数据进行对比，判断企业是否存在少报销售收入的问题。

4.1.2 数据整合方法

数据整合是将不同来源、不同格式的数据进行清洗、转换和集成，使其成为统一、规范的数据集。常用的数据整合方法包括数据清洗、数据转换和数据加载（ETL）。

4.2 风险指标构建

4.2.1 常见税收风险指标

财务指标：资产负债率 = 负债总额/资产总额，反映了企业的长期偿债能力。一般来说，工业企业的资产负债率应保持在40%-60%之间，如果资产负债率过高，可能意味着企业存在过度负债的风险，可能通过隐瞒收入或虚增成本来减轻债务压力；流动比率 = 流动资产/流动负债，反映了企业的短期偿债能力，流动比率一般应大于2；毛利率 = (销售收入 - 销售成本)/销售收入，反映了企业的盈利能力，如果企业的毛利率突然大幅下降，可能存在少报收入或多报成本的问题。

纳税申报指标：税负率 = 实际缴纳的税款/应纳税销售额，包括增值税税负率、企业所得税税负率等。如果企业的税负率明显低于同行业平均水平，可能存在偷逃税的嫌疑；增值税税负变动率 = (本期增值税税负率 - 上期增值税税负率)/上期增值税税负率，企业所得税税负变动率同理。如果税负变动率过大，可能意味着企业的生产经营情况发生了重大变化，或者存在税收违法行为。

生产经营指标：用电量、用水量、原材料消耗量等可以反映企业的生产经营规模。例如，通过建立用电量与生产产量的回归模型，可以根据企业的用电量估算其生产产量，与企业的申报销售数据进行对比，判断企业是否存在少报销售收入的问题。

4.2.2 风险指标权重确定

不同的税收风险指标对税收风险的影响程度不同，

需要确定各指标的权重。常用的权重确定方法包括层次分析法、熵权法、主成分分析法等。以层次分析法为例，首先构建层次结构模型，将税收风险评估作为目标层，财务指标、纳税申报指标和生产经营指标作为准则层，具体的风险指标作为指标层。然后通过两两比较确定各指标的相对重要性，构建判断矩阵^[3]。例如，在准则层中，认为财务指标比纳税申报指标稍微重要，纳税申报指标比生产经营指标稍微重要，根据这些比较结果构建判断矩阵。最后通过计算判断矩阵的特征值和特征向量，确定各指标的权重。经过计算，在该案例中财务指标的权重为0.5，纳税申报指标的权重为0.3，生产经营指标的权重为0.2。

4.3 模型建立与应用

4.3.1 常见分析模型

逻辑回归模型：逻辑回归模型是一种广义线性模型，用于预测事件发生的概率，适用于二分类问题。在工业企业税收疑点排查中，可以将企业分为高风险和低风险两类，使用逻辑回归模型预测企业属于高风险的概率。逻辑回归模型的公式为：

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

其中 $P(Y=1)$ 表示企业属于高风险的概率， X_1, X_2, \dots, X_n 为风险指标， $\beta_1, \beta_2, \dots, \beta_n$ 为模型的参数。

决策树模型：决策树模型是一种基于树结构进行决策的模型，通过一系列的判断条件将数据集划分为不同的子集，最终得到决策结果。决策树模型的构建过程包括特征选择、决策树生成和决策树剪枝等步骤。在工业企业税收疑点排查中，可以使用决策树模型根据企业的各项风险指标将企业分为不同的风险等级。

神经网络模型：神经网络模型是一种模拟人脑神经元结构的模型，具有强大的非线性拟合能力，适用于处理复杂的非线性问题。常见的神经网络模型包括多层感知机（MLP）、卷积神经网络（CNN）和循环神经网络（RNN）等。在工业企业税收疑点排查中，可以使用MLP模型构建税收风险评估模型，通过输入企业的各项风险指标，输出企业的税收风险评分。

4.3.2 模型应用流程

数据准备：将采集和整合后的数据划分为训练集和测试集，一般按照7:3或8:2的比例进行划分。训练集用于模型的训练，测试集用于模型的评估。

模型训练：使用训练集数据对模型进行训练，调整模型的参数，使模型能够更好地拟合数据。例如，在使用逻辑回归模型时，使用最大似然估计法估计模型的参

数;在使用决策树模型时,使用信息增益或基尼系数等指标进行特征选择和决策树生成。

模型评估:使用测试集数据对训练好的模型进行评估,计算模型的准确率、召回率、F1值等指标,评估模型的性能。 $\text{准确率} = \text{正确预测的样本数} / \text{总样本数}$, $\text{召回率} = \text{正确预测的正样本数} / \text{实际正样本数}$, $\text{F1值} = 2 * \text{准确率} * \text{召回率} / (\text{准确率} + \text{召回率})$ 。

模型应用:将训练好的模型应用到实际的税收疑点排查中,对企业的税收风险进行评估和预警。根据模型的评估结果,将企业分为高风险、中风险和低风险三类,对高风险企业进行重点监控和税务稽查。

5 应用中的优化建议

5.1 提高数据质量

建立严格数据审核机制,多层次审核企业申报数据。数据录入时设校验规则,对格式、范围、逻辑等校验,如企业销售收入数据设合理上下限,超限提示税务人员核实。审核中结合企业历史数据和行业平均水平对比分析,发现异常及时调查。制定统一数据采集标准与规范,明确数据定义、格式、采集频率等。加强与企业沟通培训,如制定工业企业电力消费数据采集标准,规定采集时间、方式与格式,确保数据准确一致^[4]。建立数据质量反馈渠道,鼓励企业和社会公众反馈问题,及时调查处理并反馈结果。对数据质量高的企业和个人给予奖励,营造良好提升氛围。

5.2 强化数据安全与隐私保护

采用加密、访问控制、防火墙等先进技术保障数据传输与存储安全。对企业敏感数据加密,授权人员才能解密访问;建立访问控制机制,设置不同用户权限,如用对称加密算法加密企业财务数据,用防火墙隔离税务内外网络。健全管理制度,明确职责流程,制定数据访问、使用、共享等安全规范,加强人员监督。定期评估修订制度,如建立数据安全审计制度,记录审计数据访问操作,及时发现处理安全事件。加强宣传教育,通过培训、讲座、宣传活动普及数据安全知识与法律法规,

提高税务和企业人员安全保密意识,如每年组织数据安全宣传周活动。

5.3 促进部门间数据共享

政府牵头组织相关部门制定统一数据标准与规范,明确数据定义、格式、编码规则等,建立数据字典统一描述定义数据元素,如制定企业基本信息数据标准,统一企业名称、信用代码、注册地址等格式编码。搭建数据共享平台,实现数据集中存储与共享交换,各部门按权限获取数据,提高效率与安全性。建立监控管理机制,实时监控管理数据共享过程,如搭建政务数据共享交换平台,实现税务、工商、银行等部门数据互联互通。健全激励机制与安全保障机制,奖励积极共享数据部门,加强安全管理,明确安全责任,签订安全协议,如建立考核机制,将数据共享纳入部门绩效考核。

结语

本文对大数据分析在工业企业税收疑点排查中的应用进行了深入研究。通过阐述大数据分析的内涵与特点,分析了工业企业税收疑点排查的重要性和传统方法的局限性。详细介绍了大数据分析在工业企业税收疑点排查中的具体应用,包括数据采集与整合、风险指标构建、模型建立与应用等方面。同时,对大数据分析在工业企业税收疑点排查应用中面临的挑战进行了探讨,并提出了相应的对策建议。未来,随着大数据技术的不断发展和完善,其在税收征管领域的应用将更加深入和广泛。

参考文献

- [1]邵天凡.大数据背景下税收征管存在的问题[J].财经界,2025,(11):156-158.
- [2]李玉姣,李连民,杨翟婷.大数据税收征管是否促进企业创新可持续性[J].税收经济研究,2025,30(01):35-45.
- [3]李腾.财税大数据在企业税收风险管理中的应用研究[J].冶金财会,2024,43(12):73-75.
- [4]宋金华.大数据背景下企业税收风险的控制策略[J].商业2.0,2024,(35):72-74.