

# 基于大数据技术的工业统计 数据质量优化方法研究

朱豫秦

(中国核电工程有限公司, 北京 100840)

**摘要** 工业统计数据是工业经济运行监测、产业政策制定的重要基础, 数据质量直接影响到工业经济治理的科学性、有效性。大数据技术的迅速发展为解决工业统计数据在采集、处理、应用等各个环节存在的质量问题提供了技术支持。本文以工业统计工作实际为立足点, 剖析当前工业统计数据质量存在的主要问题, 挖掘大数据技术和工业统计质量管控的适配性, 从技术融合、流程优化、机制建立等角度提出工业统计数据质量提升的方法, 构建工业统计数据质量全生命周期的管控体系, 实现工业统计数据质量从被动修正到主动优化的转变, 为工业统计现代化建设提供实践路径和方法参考。

**关键词** 大数据技术; 工业统计; 数据质量; 全生命周期管理

**中图分类号**: F270.7; TP274

**文献标志码**: A

**DOI**: 10.3969/j.issn.2097-3365.2026.13.025

## 0 引言

工业是国民经济的主导, 工业统计是工业经济运行分析的基础工作, 工业统计数据是否真实、准确、及时, 直接关系到产业决策的科学化、工业发展的高质量<sup>[1]</sup>。随着数字经济和工业实体经济深度融合, 工业生产经营数字化、智能化水平不断提高, 工业统计数据来源更加多元、种类更加复杂、数量呈爆发式增长, 传统统计数据处理模式和质量控制方法已经不能满足新时代工业统计工作要求。研究利用大数据技术对工业统计数据质量进行优化的方法, 成为促进工业统计体系现代化、提高工业经济治理能力的重大课题。

### 1 工业统计数据质量的核心要求

根据工业统计工作特点和大数据时代应用需求, 工业统计数据质量应该达到以下五个方面的要求。一是准确性, 数据质量的基础是数据本身可以真实反映企业生产经营活动, 数值核算准确、逻辑关系无矛盾, 保证数据客观真实性。二是完整性, 即统计数据的采集要覆盖工业生产经营的各个环节、主要指标, 不能有重要字段的缺失、时间序列的断点, 保证数据体系的全面性。三是时效性, 即数据可以实现采集、处理、上报的迅速传递, 符合工业经济变化快的特点, 给及时监测和快速决策提供数据支持。四是协同性, 即使多源、跨部门、跨企业的统计数据达到标准统一、

语义互通的目的, 消除数据壁垒和信息孤岛, 保证数据融合应用的可行性。五是可溯性, 要求统计数据全生命周期流转过程可以被记录、跟踪、核实, 确定数据的产生、处理、更改责任单位及时间点, 给质量问题追溯和责任认定提供支撑。五大核心要求相互联系、相互支撑, 精准性是基础、完整性是保证、时效性是关键、协同性是拓展、可溯性是兜底, 共同组成了新时代工业统计数据质量评价体系, 也成为开展数据质量优化工作的主要方向<sup>[2]</sup>。

## 2 工业统计数据质量的现存问题

### 2.1 数据采集环节存在标准不一、源头失准的问题

工业企业的数字化水平良莠不齐, 各个企业、不同行业使用的采集设备制式不同, 统计信息的采集接口、数据格式、指标口径等没有形成统一的标准, 某些领域存在着多头采集、重复采集的情况, 造成数源矛盾、数据重复。部分企业仍然使用人工录入方式来进行数据的收集工作, 很容易造成录入偏差、漏报和错报等不良现象的发生, 并且没有建立起可靠的设备端实时校验手段, 从而导致基础数据源头质量无法得到保证<sup>[3]</sup>。

### 2.2 数据处理环节存在技术滞后、效率偏低的问题

传统的工业统计数据处理工具是以结构化表格解析为主, 对于工业生产中出现的各种非结构化的数据如设备运行日志、视觉监测数据、音频数据等都没有

**作者简介**: 朱豫秦 (1981-), 男, 本科, 高级工程师, 研究方向: 质量安全管理。

很好的处理方法。数据清洗规则固化，校验手段单一，不能满足海量、异构、高并发的工业统计数据处理要求，造成数据清洗不到位，异常数据发现迟缓，致使数据处理效率和质量无法兼顾。

### 2.3 数据应用环节存在协同不足、价值偏低的问题

政府部门的统计机构、行业管理部门以及企业等之间缺少有效的数据共享和协调机制，各主体的数据分散存储、独立管理，从而造成信息孤岛。跨部门数据融合缺少技术支撑和制度保障，数据关联挖掘不够深入，不能发挥多源数据的协同效应，造成统计数据只能满足基础的指标核算要求，在产业趋势预测、精准政策制定等方面的支持作用没有得到充分发挥。

## 3 基于大数据技术的工业统计数据质量优化方法

### 3.1 大数据驱动的多源协同采集与实时校验

数据采集属于工业统计数据质量控制的第一个环节，借助大数据技术创建多源协同采集体系及设备端实时校验体系，从源头保证数据品质。创建统一的数据采集接口，利用大数据技术制定出统一的工业统计数据采集标准，规定数据格式、指标口径、字段要求，研发出符合各种采集设备、各种行业的智能适配中间件，使工业企业ERP系统、生产执行系统、设备监测系统各个系统的数据能够自动生成地进行采集对接，防止由于不同数据源之间格式不一样而产生的数据冲突，也减少了人工操作，降低了由于人为失误造成的错误<sup>[4]</sup>。搭建“设备一边缘一云端”三层采集和校验结构，设备端安装具有数据采集和实时校验能力的智能终端，依靠数字孪生技术搭建设备的理论运行模型，并将传感器所测得的数据同理论模型数据展开即时比较，若出现数据偏离超过设定标准的情形，则会立即发出拦截警报并实施警告，从而实现源头数据质量的提前把控。在边缘端采用边缘计算手段，针对采集来的数据实施初步清洗并且确认其时间序列的连贯情况，从而找到由于网络断开或是设备发生故障而造成的断点以及无效信息，进而执行数据预处理的工作流程。云端则承担着汇聚并存储大量的采集数据的任务，为之后深入处理和分析奠定基础。创建采集数据的动态筛选机制，采用机器学习算法智能判断采集数据是否有效，结合工业生产实际，把温湿度、气压、生产计划等有关信息综合起来，剔除极端天气、设备检修等特殊情况下无效的监测数据，提高采集数据的有效性、精准性。

### 3.2 大数据支撑的异构数据处理与异常识别

数据处理属于工业统计数据质量改善的关键部分，依靠大数据技术加强对于异构数据的处理能力，创建

智能化的数据清洗和异常识别体系，从而实现数据处理过程的质量改善。创建基于大数据的异构数据统一处理平台，将分布式存储、数据湖、数据仓库等技术融合起来，实现结构化、半结构化、非结构化工业统计数据的统一存储和规范化处理，创建解析并转化非结构化数据成可分析、可利用结构化数据的软件工具，克服传统处理工具功能上的限制。开发出融合知识图谱和动态规则引擎的智能数据清洗系统，建立包含工业统计指标、设备工艺参数、行业运行规则的工业统计知识图谱，用知识图谱对指标之间、领域之间的逻辑关系进行关联以及一致性检查，自动发现数据中存在的逻辑矛盾以及不合理的关系<sup>[5]</sup>。通过机器学习的方式从历史的质量数据中挖掘出异常数据的可能模式，并且产生相应的动态清洗规则以及校验阈值，再依靠对实时的数据流的实时更新来进行规则参数的调节工作，在保证数据清洗质量的情况下实现智能化、自适应地进行数据清洗工作。创建基于机器学习的异常数据深层次诊断模型，将随机森林、长短期记忆网络等算法融合在一起，对数据处理时找出的异常数据实施深入剖析，找出异常数据产生的原因，区分设备故障、录入失误、网络问题等种类的异常诱因并给出对应的解决建议，从而达到精确地诊断和迅速修正异常数据的目的，改进数据处理质量和效率。

### 3.3 大数据赋能的跨域协同共享与价值挖掘

数据应用是工业统计数据质量的最终反映，依靠大数据技术冲破数据壁垒，创建安全可控的跨域数据协同共享体系以及深入的价值挖掘体系，从而优化数据应用的质量和价值。搭建以隐私计算为基础的跨部门数据协同共享平台，用多方安全计算、联邦学习等隐私计算技术来保证政府统计部门、税务、电力、工信等行业主管部门以及工业企业数据的安全和隐私，并且促进多源跨域数据的协同融合。平台提前设定产值和用电弹性系数、增值税与营业收入比对这些主要校验规则，用多源数据关联来达到对统计数据相互校验的目的，从而保证数据的准确性、可靠性。创建工业统计数据深度价值挖掘模型，借助大数据关联挖掘、趋势预测等技术，剖析融合后的多源数据，探寻工业生产经营指标间的隐秘联系，从而实现工业经济运行趋向的准确预估，针对产业政策推行成效展开动态评判，促使工业统计数据由核算型转向决策支撑型，优化数据的应用水平。创建数据应用的反馈改良体系，凭借大数据技术抓取统计数据在各个应用场景下的使用反馈，剖析数据在决策支持、趋势预估等场合里的欠缺，形成数据品质改良的需求清单，反向引领数据采集、处理环

节的标准修订和流程改善,达到数据采集、处理、应用的闭环改良,加强数据与应用需求的契合程度<sup>[6]</sup>。

#### 4 工业统计数据质量优化方法实施的保障措施

##### 4.1 完善制度保障,明晰权责边界

制度保障是保证优化方法落实的基础,必须从制度上对工业统计数据质量管理工作做出明确的规定,界定各个主体的职责范围。推进工业统计质量管理制度的更新和健全,针对大数据技术的应用特点,充实数据全生命周期追踪、跨部门数据互相配合共享、数据品质分层管理等有关内容,把大数据技术支撑的质量改良办法纳入制度化、规范化轨道上来。创建企业数据质量分级认证制度,按照企业数据采集规范度、处理流程合规度、应用反馈时效度这些指标,把企业分成不同的等级,对优质的加以统计报表简化、项目审批优先安排等扶持举措,对数据存在明显问题的企业实施重点监管和指导,建立激励和制约相交织的制度架构。创建跨部门的协同工作机制,明确各个部门对于工业统计数据采集、共享、质控等各个环节的责任和分工,创建常态化的沟通协调和工作联动机制,推进多源数据协同共享和质量控制工作,破除跨部门协同的制度性梗阻。健全数据质量责任追究制度,规定企业法定代表人对统计数据真实性负主要责任,规定统计部门、数据处理机构等有关方面对于数据的真实性负有管理责任,对于存在数据造假、漏报错报或者整改不到位等情况,要从严追究相关责任人的责任。

##### 4.2 强化技术支撑,夯实硬件基础

不断加强大数据技术在工业统计领域的应用支撑,夯实硬件和数据处理技术平台基础。推进工业统计大数据平台的建立和升级,加大省级、市级统计机构对于数据采集、处理、存储、分析、质量控制等各方面的数字化基础设施建设的投入,使统计系统同企业、行业主管部门的数据系统无缝对接,提高平台对于海量数据处理、实时分析以及安全防护的能力。创建研发标准化的工业统计大数据工具包,把数据清洗、异常辨别、质量评判、趋向预估等关键功能囊括其中,研制契合不同行业、各样统计场合的模块化工具组件,从而达成工具包的标准化、轻量化推广目的。创建数据安全保障技术体系,依靠区块链、加密算法、访问控制等技术创建全面的数据安全保障体系,对工业统计数据的收集、保存、传递、使用全过程展开全方位的保护工作,从而保证数据的安全性与隐私性,并且防备数据被泄露或者被篡改的情况出现。

##### 4.3 培育人才队伍,提升应用能力

培育具有工业统计专业知识以及大数据技术运用能力的复合型人才,从而提高整个行业的数据应用能力。创建统计系统的大数据人才培育体系,制订具有指向性的培养方案。就目前的统计工作人员展开工业统计专业知识与大数据技术的双项培训,涵盖分布式存储、机器学习、数据可视化等核心技术,加强其大数据技术应用水平和质量把控能力。创建统计分析师和数据工程师双重认证体系,把大数据技术的应用能力纳入统计人员的考核和晋升体系,从而调动专业人员学习和运用技术的积极性。加强工业企业统计人才的培养和指导,统计部门会同行业主管部门、专业培训机构,对工业企业统计人员进行免费的培训和指导,主要讲解大数据背景下工业统计数据的采集标准、处理方式以及质量控制要求,提高企业源头数据的质量控制水平。积极引进具有工业统计、大数据分析、计算机技术等多种知识的高素质复合型人才到统计部门工作,给大数据技术在工业统计领域的应用和创新发展提供人才支持<sup>[7]</sup>。

#### 5 结束语

本文从大数据技术的角度出发,对基于大数据技术的工业统计数据质量优化方法进行系统的探究,立足于工业统计工作的实际情况,结合数字经济与工业实体经济深度融合的时代背景,针对目前工业统计数据质量管控所遇到的困境,探寻出一种利用大数据技术与工业统计数据质量管控深度融合的新思路和新途径,为工业统计体系的现代化建设提供参考。

#### 参考文献:

- [1] 潘慧敏,陈岩.煤炭生产企业统计数据质量控制与评估体系研究[J].煤矿现代化,2026,35(02):114-118.
- [2] 茆翠红.提升企业统计数据质量的对策研究:基于演化博弈分析[J].现代商业,2026(01):78-81.
- [3] 高芳.提高基层统计数据质量的有效策略探究[J].财经界,2025(34):56-58.
- [4] 叶哲全.房地产统计数据质量的控制标准研究[J].标准生活,2025(07):69-71.
- [5] 李作彬.影响工业统计数据质量的因素与解决措施分析[J].中国电子商情,2024(23):7-9.
- [6] 叶发明.基层工业统计数据质量管理与提升策略[J].中国商界,2024(04):151-153.
- [7] 王慧.新经济下基层工业统计数据质量的控制与提高[J].现代工业经济和信息化,2023,13(05):325-326.