

Research on the application of AI in file data quality verification and anomaly identification in the context of human resource sharing

Xiaotong Fei

China Resources Pharmaceutical Commercial Group Co., Ltd., Jinan, Shandong, 250000, China

Abstract

With the accelerated digital transformation of enterprises, human resource management is evolving towards shared service models and data integration. While centralized management of archival data has improved efficiency, challenges such as inconsistent quality, incomplete data, and frequent anomalies have emerged from multi-source data fusion. Artificial Intelligence (AI), leveraging its powerful data processing and intelligent recognition capabilities, provides innovative solutions for archival data quality verification and anomaly detection. This paper examines the current status of archival data quality management under shared service models, explores the application principles and key algorithms of AI in automated data verification, defect correction, and anomaly identification, and summarizes practical outcomes and optimization pathways through case studies. The research concludes that AI empowerment can significantly enhance archival data accuracy and security, driving intelligent and standardized high-quality development in human resource management.

Keywords

archival data; data quality verification; anomaly detection; artificial intelligence

人力资源共享场景下 AI 在档案数据质量核查与异常识别中的应用研究

费筱桐

华润医药商业集团有限公司, 中国·山东 济南 250000

摘要

随着企业数字化转型加快, 人力资源管理逐步向共享服务与数据集成模式发展, 档案数据集中管理效率提升, 但多源数据融合带来一致性差、完整性不足、异常频发等质量挑战。人工智能(AI)凭借强大数据处理与智能识别能力, 为档案数据质量核查与异常识别提供了创新手段。本文梳理了共享服务模式下档案数据质量管理现状, 探讨AI在数据自动核查、缺陷修复、异常识别中的应用原理与关键算法, 并结合案例总结实际成效与优化路径。研究认为, AI赋能可显著提升档案数据准确性与安全性, 助推人力资源管理智能化与规范化高质量发展。

关键词

档案数据; 数据质量核查; 异常识别; 人工智能

1 引言

随着信息技术的快速发展与企业管理模式的不断创新, 人力资源共享服务中心(HRSSC)在全球范围内得到广泛应用。该模式通过集中处理人力资源各项基础事务, 实现人力、财务、业务的深度集成和资源最优配置, 有效提升了服务效率和管理水平。然而, 海量的档案数据在共享服务场景下高度集中, 数据类型繁杂、结构多样, 来源分散且更新频繁, 极易引发数据冗余、重复、失真及异常等质量风险, 给

后续的人力资源决策、薪酬管理、员工发展等环节带来巨大隐患。传统的数据核查和异常识别多依赖人工抽查或规则匹配, 存在效率低、准确率有限、响应滞后等突出问题, 已难以适应现代企业高效、精准、动态的数据治理需求。

人工智能技术的崛起, 为档案数据质量管理提供了全新动力。AI通过机器学习、深度学习、自然语言处理等方法, 能够在大规模、多源、异构的数据环境下自动识别错误、填补缺失、发现模式与异常, 大幅提升数据核查的自动化与智能化水平。本文在分析人力资源共享服务环境下档案数据特点与管理痛点的基础上, 系统梳理AI在数据质量核查与异常识别领域的核心技术及应用路径, 结合典型案例, 总结经验与启示, 提出进一步完善智能数据治理体系的思考。研究

【作者简介】费筱桐(1989-), 女, 中国山东济南人, 硕士, 馆员, 从事档案研究。

旨在为企业人力资源管理数字化转型和高质量发展提供理论参考与实践借鉴。

2 人力资源共享场景下档案数据质量管理现状

2.1 人力资源共享服务的业务模式与数据特点

人力资源共享服务模式以标准化、流程化和集中化为核心，全面整合了包括人事档案管理、薪酬核算、考勤管理、社会保险福利、招聘录用等多个业务板块。在这一体系下，业务流程覆盖众多部门、分支机构以及外部合作伙伴，数据来源极为广泛，涵盖结构化的系统数据、表格，以及大量半结构化的表单、文本和非结构化的扫描件、附件等信息类型。随着数据在不同环节间的流转、归档和更新，呈现出数据格式多样、结构松散、标准不统一和实时性要求高等鲜明特征。这种多源、异构、动态的数据格局极大增加了数据整合与比对难度，同时也使档案数据质量控制和监测变得复杂。传统管理手段通常依赖人工检查或静态规则，无法满足全流程、全要素的高效精准治理需求，难以及时发现数据异常、重复或缺失等质量问题，进而影响人力资源管理的效率与决策的科学性。因此，推动智能化、自动化的质量管控体系成为提升共享服务数据治理能力的关键。

2.2 档案数据质量问题的主要类型与风险

在人力资源共享场景下，档案数据质量问题主要体现在以下几个方面：一是数据一致性差，同一员工的基本信息在不同系统或环节存在冲突或不统一现象；二是数据完整性不足，部分字段缺失或录入不全，影响数据后续分析利用；三是冗余与重复数据大量存在，导致数据资源浪费与查询效率低下；四是数据时效性差，更新不及时，信息滞后影响决策准确性；五是异常值与错误数据混杂，如身份证号、工号、工龄等关键字段异常，影响薪酬、考核等业务的合规运行。上述问题如不及时发现与修正，易引发管理风险、政策合规性风险以及员工权益争议等一系列负面后果。

2.3 传统数据质量核查与异常识别方式的局限

传统的数据质量核查主要依赖规则校验、人工抽查和系统预警等方式。人工核查因受限于人力成本和操作主观性，难以应对海量数据场景，且抽查比例低，存在遗漏和误判风险。基于规则的校验虽然能够实现部分自动化，但规则覆盖范围有限，面对复杂的数据交叉关系和潜在异常难以有效识别，且维护与更新成本高。对于非结构化和半结构化档案信息，如员工简历、合同附件等，传统工具更是无能为力。数据异常一旦出现，往往难以及时发现和处理，影响企业运转效率和管理质量。因此，亟须引入更加智能、自动化的技术手段，提升档案数据质量核查与异常识别的科学性和实效性。

3 人工智能技术在档案数据质量核查中的应用基础

3.1 AI 在数据质量核查领域的核心优势

人工智能在数据质量核查中的核心优势主要体现在强

大的自动学习与泛化能力。通过对大量历史数据的学习，AI 能够自动挖掘出数据间复杂的关联模式，识别出数据中隐藏的规律与异常，摆脱了传统规则系统对先验知识的依赖。AI 技术可对结构化、半结构化甚至非结构化的档案数据进行智能解析，实现从文本理解、实体识别到语义推理等多维度的数据质量评估。结合自然语言处理、知识图谱等技术，能够对员工档案中的文本、证书、合同等非标准化信息进行自动核查和比对，有效提升数据管理的广度与深度。

3.2 主流 AI 算法与工具在档案数据治理中的适配性

当前，主流 AI 算法在档案数据治理中展现出高度的适应性与智能化特征。机器学习算法如决策树、随机森林、支持向量机等，广泛应用于结构化档案数据中的缺失值填补、重复项消除与冲突数据识别，能够对数据质量问题进行高效分类和精准预测。对于员工简历、劳动合同等非结构化档案数据，深度学习模型（如卷积神经网络 CNN、循环神经网络 RNN、Transformer 等）在文本内容提取、语义理解及自动纠错方面具有显著优势，提升了文档处理的自动化和准确性。在异常检测环节，孤立森林、聚类分析、自编码器算法可有效识别孤立数据点、模式偏差和异常行为，增强对数据风险的预警能力。同时，AI 辅助的数据质量可视化工具实现了数据流转各环节的动态监控、问题追溯和自动修复，帮助管理者及时发现并解决数据异常，为档案数据治理和决策提供了科学、直观的支持，有力保障了人力资源管理的合规性与高效性。

3.3 AI 驱动档案数据质量核查的实施流程

AI 驱动的档案数据质量核查一般分为数据预处理、特征工程、模型训练与评估、自动核查和异常输出等环节。首先，通过自动化脚本对多源档案数据进行标准化、清洗与格式转换，确保后续分析的基础一致。其次，结合业务需求和数据特性，自动生成多维度特征，输入 AI 模型进行训练和预测。模型评估阶段，结合业务专家经验和历史校验结果优化模型参数，提升核查准确率。正式上线后，AI 系统可实现对新流入档案数据的自动质量核查与异常报警，并支持人工反馈校正机制，实现人机协同的数据治理闭环。

4 AI 在档案数据异常识别与质量提升中的应用实践

4.1 智能规则挖掘与自适应异常识别

AI 可通过机器学习与模式识别，对历史档案数据进行智能规则挖掘，自动发现潜在的业务逻辑约束和异常模式。例如，员工年龄、入职年限、学历、岗位等多维度信息之间具有明确的逻辑关联，AI 可自主学习这些关系，对不符合常规的数据自动标记并发出预警。基于自适应阈值与动态聚类的方法，可灵活调整异常检测敏感度，适应不同数据环境和业务需求。对于新出现的数据类型和业务场景，AI 模型还可持续自我优化，增强对未知异常的发现能力，显著提升

数据核查的前瞻性与准确性。

4.2 自然语言处理在非结构化档案数据核查中的应用

档案数据中大量信息以文本、图片、文档等非结构化形式存在，如员工简历、培训证明、合同附件等。自然语言处理（NLP）技术能够自动提取文本实体、理解语义内容，实现关键字段抽取、证据一致性比对及虚假信息检测。例如，通过命名实体识别（NER）、文本相似度比对、语法分析等算法，AI系统可自动审核员工简历与岗位描述的匹配度，发现学历、工作经历等虚假或异常信息。图像识别与OCR（光学字符识别）可用于扫描件、照片的文字信息提取与核查，为档案数据的全方位质量监控提供技术支撑。

4.3 智能数据修复与自动化数据清洗

AI技术在档案数据治理中的作用不仅体现在异常发现上，更在于通过智能补全与自动修复等手段实现数据质量的全面优化。利用预测模型和业务规则，AI可以针对缺失或不完整的档案字段自动生成合理的填充值，或依据历史数据和业务逻辑提出最优修正建议。面对来自不同系统的冗余和重复数据，AI能够借助聚类分析和关联规则自动完成去重和标准化，提升数据一致性。通过构建自动化的数据清洗流水线，AI实现了对海量档案数据的批量处理，有效提高了整理效率和准确率，大幅减少人工介入的时间与成本。更为重要的是，AI系统具备持续学习和自我优化能力，能够不断根据新出现的异常类型、业务需求和反馈信息迭代优化自身模型，使档案数据治理逐步实现从被动修复问题向主动识别、防控风险转变，整体提升了人力资源数据管理的科学化、智能化水平。

5 AI 赋能档案数据质量治理的成效与挑战

5.1 AI 提升档案数据质量与管理效率的实际成效

随着AI在档案数据质量核查与异常识别中的应用逐步深化，许多人力资源共享服务中心已经取得了显著成效。数据一致性与完整性水平明显提升，错误率、遗漏率持续下降，数据管理效率显著增强。自动核查与智能预警机制的实施，大幅缩短了数据问题的响应时间，减轻了一线运维人员的工作压力。AI推动下的档案数据治理体系更加规范、智能和高效，为企业高质量的人力资源管理、人才发展决策和合规风险管控提供了坚实的数据基础。

5.2 AI 应用中面临的技术与管理挑战

尽管AI带来了档案数据治理的重大变革，但在实际应用过程中仍面临诸多挑战。数据异构性与业务多样性导致模型训练和算法适配难度较大，部分场景下AI系统对复杂异常的识别精度有待提升。数据隐私与安全问题成为智能核查

过程中必须重点防范的风险，需强化数据脱敏、权限管控与合规管理。此外，AI系统的应用和维护需要高水平的数据工程和算法人才支持，部分企业在智能化转型过程中面临人员能力与系统集成双重压力。智能核查与人工审核如何高效协同，避免误判与漏判，也对企业管理能力提出了更高要求。

5.3 未来发展趋势与应对策略

展望未来，AI赋能档案数据治理将持续向深度与广度拓展。一方面，随着大数据、区块链、联邦学习等前沿技术的融合应用，档案数据的质量管理将实现更高水平的智能协同与安全共享；另一方面，企业需进一步加强数据治理体系建设，完善AI算法适配、数据标准制定与全流程管理机制，推动“数据—模型—业务”一体化发展。政策层面可推动行业标准与合规要求的统一，促进数据开放与智能治理创新协作。通过技术创新与管理优化协同发力，有望实现人力资源共享场景下档案数据质量管理的持续突破与价值创造。

6 结语

人力资源共享服务的快速发展对档案数据质量和治理能力提出了更高要求。人工智能技术的深度融合，推动了档案数据核查和异常识别从规则驱动、人工主导向智能化、自动化转变，有效解决了多源异构、海量动态数据环境下的质量控制难题。实践表明，AI技术不仅提升了数据一致性、完整性与安全性，还极大增强了人力资源管理的科学性与敏捷性。面对未来数据治理的复杂形势，企业应持续加大AI应用创新力度，完善数据标准、强化人才培养，构建人机协同、高效智能的数据治理新格局。通过技术进步与管理创新的双轮驱动，人力资源管理必将在数字经济时代实现更高质量、更加智能和可持续发展。

参考文献

- [1] 卢晓洁.企业人力资源管理效能提升策略探索——基于人事档案大数据的视角[J].经济师,2025,(07):254-256.
- [2] 王丽娜,杨植林,冯军.基于大数据背景的人事档案信息化建设分析[J].黑龙江档案,2025,(03):279-281.
- [3] 刘园园.事业单位人事档案数字化管理的实施路径与优化策略[C]//江西省工程师联合会.第二届智能工程与经济建设学术研讨会论文集(二).江苏省徐州市丰县人力资源办公室;2025:459-462.
- [4] 林凌.数字化时代人力资源发展新趋势下人事档案管理路径研究[J].中国科技投资,2025,(06):110-112.
- [5] 郭栋.基于信息化手段的集团企业穿透式监管与效能提升研究[J/OL].煤炭经济研究, 1-9[2025-09-07].