

Application Scenarios and Optimization of Data Mining Technology in Archival Management

Zhichao Liang

Office of the Institutional Organization Committee of the Handan District Committee of the Communist Party of China, Handan, Hebei, 056001, China

Abstract

Against the backdrop of digital transformation, archival information is experiencing exponential growth. Traditional archival management has revealed drawbacks such as high search costs and the inability to fully exploit its value. However, the ability of data mining to uncover correlations within massive datasets can provide technical support for this challenge. This paper combines archival management practices to specifically elaborate on the application of data mining technology in scenarios such as archival classification, retrieval, and risk assessment. It also discusses issues arising during the implementation of data mining technology, including poor data quality, insufficient algorithm adaptability, and inadequate security protection systems. Corresponding optimization pathways and improvement directions are proposed to promote the practical application of data mining technology in archival management, fully unlock the value of data resources, and provide technical references for archival management work.

Keywords

Archival Management; Data Mining; Application Scenarios; Optimization Strategies; digital transformation

档案管理中数据挖掘技术的应用场景与优化

梁志超

中共邯郸市邯山区委机构编制委员会办公室机构编制研究中心, 中国·河北 邯郸 056001

摘要

在数字化转型背景下, 档案信息呈海量增长, 传统档案管理暴露出搜索成本高、无法完全挖掘利用价值等弊端, 而数据挖掘能够发现海量数据关联性的特点可以为提供技术支撑。本文结合档案管理业务, 对数据挖掘技术在档案分类整理、档案检索、风险评估等场景的应用情况进行了具体阐述, 同时论述了数据挖掘技术应用过程中出现的数据质量不佳、算法适配性不足、安全保护体系不完善等问题, 提出了相应的优化路径与改进方向, 以期推动数据挖掘技术在档案管理工作中的落地应用, 充分挖掘数据资源价值, 为档案管理工作提供技术参考。

关键词

档案管理; 数据挖掘; 应用场景; 优化策略; 数字化转型

1 引言

数字社会, 档案既记载社会发展, 也记录人类历史。档案数据正从纸质介质向电子介质发生翻天覆地的转型, 其数量、种类的变化速度极快, 以保管、存储为主的传统档案管理方法已不能适应当下对档案数据“快用、全用、深用”的需求。大数据时代呼唤数据挖掘, 在浩瀚如海的数据中寻求规律、提炼信息, 为档案管理注入新鲜的活力。将数据挖掘贯穿于档案管理的各个环节, 拓宽应用场景, 推动应用落地, 提升管理效能, 是值得管理者思考和探索实践的问题, 对推动档案事业高质量发展、发挥档案资政服务与经济

价值具有重要意义。

2 数据挖掘技术在档案管理中的核心应用场景

2.1 档案智能分类与标准化整理

档案分类存在门类繁多、内容繁琐、人工分类整理效率低、出错率高、内容无法统一归并等问题, 数据挖掘中的聚类分析算法可根据档案内容、形成时间、主题等信息数据, 自动识别档案间的相似度, 对档案进行聚类分类; 依据档案分类模型, 输入档案名称、正文内容、格式类型等基础信息, 通过上述聚类分析算法将文书档案、科技档案、声像档案等按类别分类, 并根据档案内容的相似度区分不同类别的子类^[1]。智能分类不仅节省人工成本与档案整理归类时间, 还能避免人为错误导致的档案归类偏差, 使档案整理更趋规范化、标准化, 为后续档案管理工作奠定坚实基础。

【作者简介】梁志超(1984-), 男, 中国河北邯郸人, 助理档案管理员, 从事档案管理研究。

2.2 档案信息精准检索与关联推荐

传统检索主要通过键检索，检索范围小或检索结果重复，用户通过检索结果再进行筛选，降低利用效率。数据挖掘中的关联规则挖掘和语义分析方法可实现对档案信息的精确检索与推荐。通过档案数据挖掘挖掘文本语义、主题关联、逻辑关系等，构建档案信息知识图谱，输入检索键，系统将检索到其相关档案，通过关联规则挖掘其他与之相关的档案，建立同检索对象相关的完整信息链，通过用户过去检索信息和利用信息，利用协同过滤算法，为用户做好推荐工作，方便用户寻找信息，实现更加精确方便的检索利用。

2.3 档案风险预警与安全管控

保证数据安全完整是档案管理的基本要求。在档案数据的存储、传输与利用过程中，均存在数据外泄、篡改、丢失等风险。利用数据挖掘技术对档案管理全过程数据进行分析，并构建风险预警模型，对可能出现的风险进行预警；通过分析档案访问日志、操作日志、存储日志等数据记录，识别异常访问行为、异常操作行为及存储介质异常行为等潜在风险，并及时预警；利用数据挖掘技术对档案数据开展完整性分析、逻辑关系分析及特征值分析，及时发现数据篡改、数据丢失等问题，确保档案数据真实完整、安全可靠^[2]。通过主动风险预警实现安全可控，突破被动防御模式，为档案资源安全提供全面保障。

3 档案管理中数据挖掘技术应用面临的主要问题

3.1 档案数据质量参差不齐，基础支撑薄弱

当前，档案信息采集、录入与整合环节的弊端，对档案数据挖掘效果的影响显著。一是档案数据采集与录入环节存在缺陷，因部门、时间段不同，档案数据的格式、字段、编码存在差异，导致数据格式不兼容、信息不完整；二是传统录入环节的弊端，如录入错别字、数据重复、字段空值等；三是电子档案转换过程中的弊端，如格式紊乱、字段空值等。另外，档案数据中还包含大量非结构化数据，如扫描件、图片、音频、视频等，这些非结构化数据缺乏固定解析标准，难以直接用于数据挖掘分析。数据质量缺陷致使数据挖掘模型难以获取准确、有效的数据输入，进而影响模型训练效果及分析结果稳定性，这成为数据挖掘技术在档案管理中应用的首要障碍。

3.2 技术适配性不足，应用场景固化

目前数据挖掘技术应用于档案管理存在着技术与现实需求结合不紧，应用场景单一的问题。其一，由于目前的数据挖掘算法大多为通用算法，没有针对档案数据的特点进行优化改进，档案数据的多源、异构、时序等特点不能被通用算法完全适应，导致分析结果没有针对性和实用性。如对档案文本进行语义分析的算法，如若没有针对档案术语、行文要求等特点进行优化，会导致语义分析不准确，检索、分类

不准确。其二，数据挖掘大多应用在档案分类、档案检索等简单的场景，对于档案价值判断、趋势预测、决策支撑的应用场景较少，应用场景单一^[3]。其三，部分档案部门引进了数据挖掘软件，但操作复杂，难以与现有的档案管理系统结合使用，需要专职的技术人员维护，基层部门缺乏技术人才，技术难以落地。

3.3 安全保障体系不完善，隐私与权益风险突出

其中包含大量敏感信息，如个人隐私信息、企业商业秘密、国家秘密信息等，数据挖掘技术在挖掘信息的同时，可能伴随信息泄露（含隐私泄露）等安全隐患。在当前档案管理工作中，数据挖掘技术的安全保障措施尚未完全落实，仍存在诸多安全问题。第一，数据挖掘技术应用中，数据脱敏处理技术不完善，导致数据存在被窃取、滥用的风险。第二，档案管理系统的安全防护技术手段较为薄弱，难以抵御黑客攻击、病毒入侵等风险，数据在传输与存储过程中可能面临安全威胁。第三，数据挖掘技术应用中缺乏有效的安全管理措施与责任机制，对数据挖掘过程中的数据访问权限、使用流程等缺乏有效监管，易出现内部人员窃取、滥用数据的情况，进而引发安全事故。第四，数据挖掘技术应用还可能涉及档案著作权、个人信息权益等法律问题，目前对档案数据挖掘的权利归属、使用范围等缺乏明确规定，易引发纠纷，不利于数据挖掘技术的推广应用。

4 档案管理中数据挖掘技术应用的优化路径

4.1 构建标准化数据治理体系，夯实技术应用基础

从收集、清洗、融合、规范等不同环节入手，建立全流程数据管理机制。首先，制定档案数据收集标准规范，明确各项档案数据的收集范围、字段、格式、编码等标准，规范档案数据收集行为。推进纸质档案数字化转型，采用扫描、OCR（光学字符识别）等技术提高数据录入准确率，同时建立数据录入审核机制，安排专人对录入数据进行审核校验，及时纠正错误数据。其次，加强档案数据清洗与融合，运用数据清洗工具过滤重复数据、缺漏数据、错误数据，通过数据转换技术将其转换为统一格式，实现数据来源与数据类型的融合。针对非结构化数据，采用自然语言处理、图像识别、音视频识别等技术提取所需信息，将其转换为结构化数据，提升数据挖掘价值。其三，构建档案数据质量评估机制，明确数据完整性、准确性、一致性、时效性等评价标准，定期对档案数据质量进行评估与改进，为数据挖掘技术的应用提供数据质量保障^[4]。

4.2 推动技术适配性优化，拓展多元化应用场景

从算法、工具、场景三个层面提升技术适配性与应用广度。一是加强个性化算法研究，针对档案数据本体的多样性、异构性、时序性等特点，对现有通用算法进行改进，例如将档案术语字典、档案文理逻辑规则融入文本挖掘算法，提升算法精度；针对档案数据本体的时序化特点，运用时序

数据挖掘算法挖掘档案数据的时间规律性与趋势性。加强跨界融合,联合高校、科研机构、企业等共同攻关,研发档案数据专有算法。二是研发智能、易用且兼容现有档案管理系统工具,通过精化操作流程、制作可视化操作界面,降低基层档案管理人员的操作难度。打造模块化工具体系,支持灵活调用。三是拓宽数据挖掘技术的应用领域,由档案管理领域向决策服务领域延伸。在档案价值鉴定领域,通过挖掘档案的利用频率、相关度、社会影响度等数据,建立档案价值鉴定模型,为档案的留存、销毁、利用提供依据;在趋势预测领域,通过挖掘档案历史数据预测未来趋势,为决策提供依据;在政务服务领域,通过挖掘民生领域档案数据简化业务流程,为政务服务提供参考依据。

4.3 完善安全保障与合规体系,防范各类风险

要确保数据挖掘技术在档案管理中合法应用,必须建立起“技术+制度+法律”三位一体的安全保障体系。从技术上,需建立档案管理系统安全防护体系:借助加密技术对档案数据进行加密传输与加密存储,防止数据泄露;借助访问控制技术实施数据角色权限管理,明确不同人员的访问权限与操作范围,严格控制数据访问行为;借助安全审计技术记录数据挖掘过程日志,对数据异常行为及时进行处置。建立数据脱敏机制,对档案中的敏感信息进行脱敏处理,既要满足数据挖掘需求,也要保障敏感信息安全^[9]。建章立制方面,需建立数据挖掘安全管理规范,规范并明确数据挖掘的操作流程、安全职责、风险处置流程;建立数据质量与数据安全考核机制,将数据质量与数据安全纳入档案主管部门绩效考核范围,提升工作人员的安全意识与责任意识。在法制层面,需加强档案数据挖掘相关法律法规研究,明确档案数据权利归属、使用边界及隐私保护等法律规范;档案数据管理部门在开展数据挖掘工作时,需严格遵守《中华人民共和国档案法》《中华人民共和国个人信息保护法》等法律法规,确保技术应用合法合规;加强对档案管理人员的法律知识培训,增强其法律意识与风险意识,防范因操作不当引发法律风险。

4.4 强化人才队伍建设,提升技术应用能力

人才是数据挖掘技术在档案管理领域应用的重要保障,

培养一支既熟悉档案管理知识,又掌握数据挖掘技术的复合型人才队伍。一方面加强对档案管理人员的技术培训,建立档案管理人员培训体系,进行数据挖掘基础知识、常用算法、常用工具、安全防范等培训,线上线下培训相结合,提高管理人员技术应用能力,鼓励档案管理人员进行行业交流、学术研讨,学习先进的实践经验、技术和理念。另一方面建立人才引进机制,档案管理部门可通过招聘,进校园招聘,社会招聘等途径招引进数据科学、计算机技术、信息管理等相关专业的高水平技术人才加入人才队伍。建立人才激励机制,对在数据挖掘技术应用和创新研究方面取得突出成绩和贡献的人才,要给予表彰和奖励,激发人才的工作热情和创造力。通过人才队伍建设,为档案工作中数据挖掘技术的广泛应用提供人才资源。

5 结语

数据挖掘技术是数字化和智能化档案管理的核心技术支撑,在档案目录管理、数据检索、风险评估等领域将获得广泛应用。但数据挖掘技术在档案管理应用中仍存在数据质量、技术适配、安全防护等核心问题,制约其作用充分发挥。健全数据治理标准、提升技术适配性、加强数据安全与合规管理、强化人才队伍建设,是实现数据挖掘技术应用突破的关键路径,可推动数据挖掘技术与档案管理深度融合。在新的技术形势下,数据挖掘技术在档案管理应用中将继续普及,为档案管理的创新发展提供有力支撑,助力数字化时代档案管理高质量发展。

参考文献

- [1] 蒙泓好.数据挖掘技术在档案管理系统中的应用研究[J].中国新通信,2025,27(18):39-41.
- [2] 王涛.基于数据挖掘技术的事业单位档案管理信息优化研究[J].中原文化与旅游,2025,(15):177-179.
- [3] 张越,宋文霞,刘雪菲,等.基于数据挖掘技术的计量科研档案管理创新范式初探[J].中国计量,2025,(07):80-84+95.
- [4] 黄伟红.基于数据挖掘与智能分析的高校档案管理体系构建分析[J].兰台内外,2025,(06):75-77.
- [5] 王琴钗.数据挖掘技术在档案管理中的应用[J].集成电路应用,2025,42(01):296-297.