

Research on the application path of AI technology in intelligent classification and retrieval of archives

Chan Meng

Yijinhuluo Banner Human Resources Comprehensive Service Center, Ordos, Inner Mongolia, 017200, China

Abstract

The rapid advancement of information technology, particularly the continuous progress in artificial intelligence (AI) systems, has brought transformative changes to the archival management industry. As a critical task in archival administration, intelligent classification and retrieval of archives directly impacts management efficiency and quality. Traditional manual methods for processing massive data are cumbersome and inefficient, failing to meet modern society's demands for efficient and precise archival management. However, AI technologies such as machine learning, natural language processing, and image recognition can automate archival classification and retrieval through smart algorithms, significantly enhancing both efficiency and accuracy. This paper explores application pathways for AI in intelligent archival classification and retrieval, analyzes its current status, advantages, and challenges, while proposing optimization suggestions. Research demonstrates that implementing AI technology elevates the automation level of archival management, drives intelligent development, and establishes a robust technical foundation for digital archival management.

Keywords

artificial intelligence; archives management; intelligent classification; intelligent retrieval; machine learning; natural language processing

AI 技术在档案智能分类与检索中的应用路径研究

梦婵

伊金霍洛旗人力资源综合服务中心, 中国·内蒙古 鄂尔多斯 017200

摘要

信息技术迅猛发展, 尤其是人工智能 (AI) 技术持续进步, 给档案管理行业带来巨大变革。档案智能分类与检索作为档案管理的重要任务, 关乎管理效率与质量。传统人工方式处理海量数据时, 繁琐且效率低, 难以满足现代社会对档案管理高效、精准的要求。而AI技术中的机器学习、自然语言处理和图像识别等, 可借助智能算法实现档案分类与检索自动化, 大幅提升效率和精准度。本文探讨AI技术在档案智能分类与检索中的应用路径, 剖析其应用现状、优势与挑战, 并提出优化建议。研究显示, AI技术引入能提升档案管理自动化程度, 推动其智能化发展, 为数字化档案管理筑牢技术根基。]

关键词

人工智能; 档案管理; 智能分类; 智能检索; 机器学习; 自然语言处理

1 引言

档案作为重要的信息载体, 承载了大量的历史记录和文化财富。随着信息时代的到来, 档案的存储与管理方式也发生了深刻的变化。传统的档案管理方式大多依赖人工分类、整理和检索, 这种方式不仅效率低下, 而且容易产生人工疏漏和误差, 导致档案管理的质量无法满足快速发展的社会需求。随着数字化档案的逐步推广和应用, 传统的人工处理方式逐渐暴露出无法应对海量信息的缺陷, 档案管理领域急需引入更为高效、智能的技术。

人工智能 (AI) 技术, 作为当今最具前景的技术之一,

【作者简介】梦婵 (1989—), 女, 蒙古族, 中国内蒙古科右前旗人, 本科, 馆员, 从事干部人事档案研究。

已在多个领域取得了显著的应用成效。在档案管理中, AI技术的引入为档案的智能分类与检索提供了新的解决方案。通过利用机器学习、自然语言处理 (NLP)、图像识别等AI技术, 档案管理系统能够自动识别和分类各种档案数据, 并根据用户的需求进行高效的检索, 极大提高了档案管理工作效率和精准度。智能化的档案管理系统不仅能够减轻人工工作负担, 还能在大数据时代满足快速、高效的信息管理需求。

2 档案智能分类与检索的现状与挑战

2.1 传统档案分类与检索方法的局限性

传统的档案分类与检索方法多依赖于人工操作。人工分类通常基于文件内容的主题、格式等进行分类, 然而在面对庞大的档案量时, 人工分类不仅耗时, 而且容易出现重复、

错误或遗漏的情况^[1]。同时,人工检索则通常依赖于关键词或特定的索引方式,但这种检索方式在面对信息量巨大的数字档案时显得非常低效。档案检索的准确性和实时性也受到许多因素的影响,如检索人员的经验、输入的关键词等。

在数字化档案管理过程中,传统的分类与检索方法更加无法满足大数据环境下的高效需求。随着档案数据量的不断增加,人工管理的负担日益加重,导致档案管理效率低下、准确性下降。因此,如何突破传统方法的局限,提升档案分类与检索的效率和准确性,成为当前档案管理领域亟待解决的重要问题。

2.2 AI 技术的应用现状

随着人工智能技术的发展,AI在档案管理领域的应用逐渐得到关注。AI技术,特别是机器学习、自然语言处理(NLP)和图像识别技术,为档案智能分类和检索提供了强有力的技术支持。目前,AI技术在档案管理中的应用已经取得了一定的进展,尤其在智能分类和自动化检索方面,已经有了实际的应用案例。例如,通过机器学习算法,AI可以自动从大量的文档中提取关键信息,进行分类和标注,实现自动分类的功能^[2]。同时,AI还可以通过自然语言处理技术分析文档内容,自动识别文档的主题,帮助用户进行精准的检索。

尽管AI技术在档案管理中的应用取得了一定的成果,但仍然面临诸多挑战。例如,档案数据的标准化问题仍未得到有效解决,许多档案文件在格式和内容上存在较大的差异,如何处理这些非标准化的数据是AI技术应用中的一个难点。此外,AI系统的训练数据和模型的优化问题也是实现智能分类与检索的关键。为了提高AI技术在档案管理中的应用效果,还需要解决数据标注、模型训练等方面的技术难题。

2.3 AI 技术在档案管理中的挑战

尽管AI技术在档案智能分类与检索中具有巨大的潜力,但在实际应用中仍然面临不少挑战。首先,数据的质量和标准化问题是AI技术能否成功应用的关键。档案数据通常包含大量的非结构化信息,如扫描件、手写文档等,这些数据的处理和转化对AI系统提出了较高的要求^[3]。其次,AI技术的训练和优化过程需要大量的标注数据,这对于许多档案管理机构来说可能是一个巨大的挑战。最后,AI技术的复杂性和模型的可解释性问题也是目前应用中的一个瓶颈,尤其是在对分类和检索结果进行解释时,如何确保AI系统的决策过程透明且符合实际需求,仍然是一个亟待解决的问题。

3 AI 技术在档案智能分类与检索中的应用路径

3.1 机器学习在档案智能分类中的应用

机器学习是AI技术中最为广泛应用的领域之一,其通过算法模型对大量数据进行训练,从中提取有价值的信息,并利用这些信息进行预测或决策。在档案智能分类中,机器

学习可以通过对档案内容进行特征提取,实现自动分类。具体而言,机器学习可以通过监督学习、无监督学习等方法,对档案数据进行训练,形成分类模型,从而使AI系统能够根据档案的内容和标签进行自动分类。

在实际应用中,机器学习技术可以被用来对档案进行主题建模,根据文档的内容自动识别并归类为不同的主题类别。例如,利用支持向量机(SVM)、随机森林(Random Forest)等算法,可以有效地对档案文献进行分类,提高分类的准确性和效率。同时,深度学习技术,尤其是卷积神经网络(CNN)和长短时记忆网络(LSTM)等算法,也在处理复杂的非结构化数据中表现出色,为档案管理系统提供了强大的分类能力。

3.2 自然语言处理在档案智能检索中的应用

自然语言处理(NLP)是AI技术中的另一个重要领域,特别在文本数据的处理和理解方面具有重要应用。在档案智能检索中,NLP技术能够帮助AI系统理解档案中的文本信息,从而进行更加精准的检索。NLP技术可以通过文本挖掘、词向量表示、实体识别等技术,对档案数据进行语义理解,提升检索的准确性。

例如,基于自然语言处理的搜索引擎可以通过对用户输入的查询进行语义理解,结合档案内容进行精准匹配,提供相关的检索结果。通过词嵌入技术(如Word2Vec、BERT等),AI系统能够识别查询词和档案内容之间的语义关联,自动理解同义词、上下文信息等,实现更高效、更精准的检索^[4]。此外,语义搜索技术可以帮助用户不仅根据关键词,还可以根据意图和语境来进行信息检索,显著提高档案信息的可获取性。

3.3 图像识别技术在档案分类与检索中的应用

图像识别技术在档案管理中也有着重要的应用,尤其在处理扫描档案和图像数据时。通过图像识别技术,AI系统能够自动识别和分类图像中的信息,并将其转换为可检索的文本或标签。例如,针对扫描的档案图像,AI可以通过OCR(光学字符识别)技术提取其中的文字信息,进行后续的分类和检索。同时,图像识别技术还可以应用于照片档案、医疗影像档案等的分类管理,帮助自动标注和检索图像数据中的关键信息。

图像识别技术还能够与其他AI技术结合,提高检索的精准度。例如,在档案管理系统中,AI可以根据图像的内容进行相似度检索,帮助用户快速找到相关的档案资料。通过图像数据的自动化处理和分析,图像识别技术为档案管理提供了更加全面的智能支持。

4 AI 技术在档案智能分类与检索中的挑战与发展前景

4.1 技术难题与挑战

虽然AI技术在档案管理中展现了巨大的潜力,能够有效提高档案分类与检索的效率和精准度,但在实际应用过程

中,仍然面临着一系列技术难题。首先,数据质量和标准化问题仍然是 AI 应用的瓶颈之一。档案数据通常由各种来源组成,其中包括扫描件、手写文本、照片、音频和视频等多种形式的非结构化数据。如何处理这些非结构化数据,提取有用信息,成为 AI 技术在档案管理中面临的重要挑战。对于手写文本和扫描件,尽管光学字符识别(OCR)技术已经取得了一定的进展,但在面对模糊、失真或者复杂布局的文档时,OCR 技术的识别率依然存在较大局限。

其次, AI 算法的可解释性问题也是当前 AI 技术应用中的挑战之一。现代 AI 技术,尤其是深度学习模型,通常被认为是“黑箱”,其内部决策过程往往难以理解和解释。这种缺乏透明度的问题,在档案管理的应用中尤为突出,因为档案信息的准确性和分类结果的可追溯性直接影响到后续的决策和操作。如果 AI 系统的决策过程不透明,管理人员就难以信任其结果,甚至可能会出现错误的分类和检索结果,这对档案管理系统的有效性和安全性构成了威胁。

4.2 发展前景与优化建议

随着 AI 技术的不断进步,特别是深度学习、自然语言处理和图像识别等技术的发展, AI 在档案智能分类和检索中的应用将持续深化,展现出巨大的应用潜力和前景。未来, AI 技术将在数据处理、语义分析、图像识别等方面取得更大的突破,从而进一步提高档案管理的智能化水平,尤其是在大规模档案管理和自动化操作方面, AI 技术将为档案管理提供更强有力的支持。通过智能化的分类和检索, AI 能够有效降低人工干预,提高档案管理的效率,并为用户提供更精准的服务。

为了进一步促进 AI 技术在档案管理中的应用,相关研究机构和技术公司应加大在数据标准化、模型优化和算法可解释性方面的研究力度。数据标准化是 AI 系统能够高效工作的前提,只有确保档案数据在结构和格式上的一致性, AI 模型才能够顺利进行训练和预测。因此,档案管理领域需要推动档案数据标准化工作,加强与数据清洗和标注工具的结合,使得数据能够被 AI 系统更好地理解 and 处理。

此外, AI 算法的可解释性问题是制约其在档案管理领域广泛应用的另一个重要因素。为了使 AI 系统在档案管理中的应用更加可信和有效,未来的 AI 技术应注重算法的可

解释性。通过研发可解释的 AI 模型,能够使管理人员了解 AI 的决策过程,帮助其做出更明智的判断。

4.3 技术创新与跨领域合作的推进

为了推动 AI 技术在档案智能分类与检索中的进一步应用,技术创新和跨领域的合作将是不可忽视的关键因素^[5]。 AI 技术的成功应用需要跨领域的合作,例如档案管理领域的专业人员、技术开发人员、数据科学家和法律专家等多方协作。档案管理系统不仅要具备高效的数据处理能力,还应确保其合规性和安全性。因此,未来的技术创新将不仅仅局限于算法优化,还需要在隐私保护、数据加密、信息安全等方面进行创新。

5 结语

AI 技术的引入为档案智能分类与检索提供了新的解决方案,显著提升了档案管理的效率和准确性。通过机器学习、自然语言处理和图像识别等技术,档案管理系统能够实现自动化的分类和检索,为用户提供更加快速和精确的服务。尽管 AI 技术在档案管理中已经取得了显著成效,但在数据标准化、算法优化、可解释性等方面仍面临诸多挑战。随着技术的不断发展, AI 将在档案管理中发挥越来越重要的作用,推动档案管理行业向智能化、数字化方向发展。

未来,随着 AI 技术的不断创新,档案管理将进一步智能化, AI 系统将不仅能提高档案分类和检索的效率,还能为档案管理的整体流程提供更加智能化的解决方案。通过技术创新和跨领域合作, AI 技术在档案管理中的应用前景广阔,必将为数字化档案管理带来深远的影响。

参考文献

- [1] 姜瑞强,王雨思,袁嘉新,等.智能档案开放审核系统构建研究——以北京市朝阳区档案馆为例[J].北京档案,2025,(07):45-48.
- [2] 陈雨佳,钱明辉,杨峰,等.大模型技术赋能人事档案开发利用的挑战与对策[J/OL].档案管理,2025,(04):78-82[2025-08-08].
- [3] 靳晶.人工智能视角下的声像档案管理研究[J].数字通信世界,2025,(07):189-191.
- [4] 梁冰卿.美英加澳国家级档案馆中期战略规划比较研究及启示[J].中国档案,2025,(07):72-73.
- [5] 陈道,朱秋雨,明均仁.人工智能在研究生学术信息搜寻挫折与应对行为中的应用研究[J].情报探索,2025,(07):72-81.