

# 在档案信息深度开发中知识挖掘技术的应用

曹虹

湖南农业大学档案馆

DOI:10.12238/acair.v2i4.10305

**[摘要]** 本研究聚焦于档案信息深度开发中知识挖掘技术的独特贡献。明确了知识挖掘这一前沿概念的核心内涵及其背后的科学原理,接着深入剖析了其重要分支,如数据挖掘、文本挖掘和语义解析等,展示了它们在应对海量档案资料挑战中的关键作用和适用场景。这些技术的应用,如自动化的档案分类,显著提升了编目精度和速度;同时,通过对档案内容的深度剖析,挖掘出了隐藏的知识,极大地推动了档案信息的深度理解和增值利用。在档案管理实践中,知识挖掘技术被巧妙地融入,从而实现了流程的智能化升级和效率的显著提升。最后,详细阐述了将知识挖掘技术应用于档案信息深度开发的实践路径,包括前期的数据预处理、模型构建、信息深度挖掘以及知识的可视化表达,以期为档案管理领域提供全面的策略支持和理论框架。

**[关键词]** 知识挖掘; 数据分析; 档案; 信息处理; 技术应用

**中图分类号:** G270 **文献标识码:** A

## The application of knowledge mining technology in the in-depth development of archival information

Hong Cao

Archives of Hunan Agricultural University

**[Abstract]** This study focuses on the unique contribution of knowledge mining technology in the in-depth development of archival information. This paper clarifies the core connotation of the cutting-edge concept of knowledge mining and the scientific principles behind it, and then analyzes its main branches, such as data mining, text mining, and semantic parsing, and shows their key roles and application scenarios in coping with the challenges of massive archival materials. The application of these technologies, such as automated archival classification, has significantly improved the accuracy and speed of cataloging; At the same time, through the in-depth analysis of the content of the archives, the hidden knowledge points are excavated, which greatly promotes the in-depth understanding and value-added utilization of the archive information. In the practice of archives management, knowledge mining technology is skillfully integrated, so as to achieve the intelligent upgrade of the process and the significant improvement of efficiency. Finally, the practical path of applying knowledge mining technology to the in-depth development of archival information is elaborated, including data preprocessing, model construction, in-depth information mining and visual expression of knowledge in the early stage, in order to provide comprehensive strategic support and theoretical framework for the field of archival management.

**[Key words]** knowledge mining; Data analysis; Archives; Information processing; Technology application

### 引言

随着科技的日新月异,档案资料的规模与多样性呈现出爆炸式增长,这对档案管理的有效性提出了严峻考验。在这个背景下,知识挖掘技术作为一种前沿的数据处理策略,其潜力在于从庞大的档案资源中萃取精华信息,推动档案管理迈向智能与精细的新高度。本研究聚焦于探究知识挖掘技术在深挖档案信息

价值的应用路径,将深入剖析这一技术的核心理念、运作机制、显著优势以及实际应用场景,同时提出一套切实可行的实施框架,为档案信息管理领域的创新实践提供富有洞察力的策略和途径。

### 1 知识挖掘技术概述

#### 1.1 知识挖掘的概念与原理

知识挖掘是一种前沿的技术,它涉及对大量数据的深入分析,以揭示隐藏的结构、关联规则以及其他潜在的有益洞察。这项技术的核心概念是运用数学、统计分析、机器智能和数据库管理,从海量的数据集中筛选出有助于决策的知识。通常,知识探索的过程涵盖了数据前期处理、模式识别和结果解读等关键阶段。在数据预处理环节,数据经过清洗、转化和压缩,旨在提升数据的准确性和适用性;在模式识别阶段,各种算法被用来识别和解析数据中的规律和联系;通过结果的解释和可视化展示,将挖掘到的知识转化为易于理解的形式,以此作为支持决策的依据。<sup>[1]</sup>

### 1.2 常用的知识挖掘技术

知识挖掘领域的技术多样性丰富,每一种方法都有其独特的应用特性和效能。基础层面上,数据探宝术通过深度的统计洞察和智能学习策略,从严谨的数据结构中提炼出隐藏的规律和智慧。文本洞察技术则聚焦于解锁非结构化文本的宝藏,广泛应用于情绪解析、主题理解等复杂任务。网络探究技术更是独步于社交和互联网环境,深入剖析节点间互动和网络架构,揭示出隐秘的行为模式和动态关联。语义解析技术凭借自然语言处理的精湛技艺,能从字句中抽丝剥茧,萃取深层次的含义,为知识图谱的编织和智能查询系统提供强大支持。

### 1.3 知识挖掘技术的优势与适用场景

大规模数据处理和复杂模式解析中,知识挖掘技术展现出了无可替代的重要性。它的应用领域犹如一个多维度的宝库。商业环境中,知识挖掘技术作为强大的工具,被嵌入客户互动管理、深度市场洞察和精准销售预判,从而推动企业精细调整资源分配和市场策略,实现竞争优势。医学研究领域,知识挖掘技术挖掘电子健康记录和基因序列信息,揭示疾病的潜在联系,为疾病识别和定制化疗法提供了新视角。<sup>[2]</sup>金融业界,知识挖掘技术的应用如虎添翼,涵盖了风险评估、欺诈识别以及投资策略优化,显著提升了金融服务的稳健性和运营效率。

## 2 知识挖掘技术在档案信息深度开发中的应用场景

### 2.1 档案分类与编目优化

在档案管理的革新过程中,知识挖掘技术崭露头角,成为提升分类与编目效能的关键驱动力。传统的档案处理方式往往依赖人力,效率低且易犯误。知识挖掘技术则开创了全新的可能性,它运用前沿的文本解析、人工智能理解和机器学习策略,实现了档案数据处理的自动化革命。这项技术通过深入剖析文件内容,自动解析并提炼核心要素和主题,实现了智能化的分类过程。知识挖掘技术还重构了编目框架,通过对既有档案体系的深度洞察和持续优化,确保了编目的科学性和系统性。这种创新不仅极大地提升了档案管理的精确度和速度,而且显著降低了人力成本和错误率,为档案管理的数字化转型提供了强大的技术基石。

### 2.2 档案内容分析与知识提取

在档案内容分析与知识提取领域,知识挖掘技术展现了其显著的效能提升。传统的档案解析方法主要依赖人工操作,其效率低下且难以揭示隐藏于海量资料中的深层次知识。然而,借助

现代科技的力量,如自然语言处理、语义理解与人工智能算法,档案内容的深入分析得以自动化进行,从而挖掘出珍贵的信息宝藏。例如,自然语言处理技术能够精准地定位档案中的核心概念、实体和事件,而语义分析则能洞察文档的潜在意义和复杂的关联结构。<sup>[3]</sup>档案管理员得以迅速掌握关键信息,揭示档案间的内在联系,极大地提高了档案信息的实用性和可利用价值。

### 2.3 档案管理流程的改进与优化

在档案管理工作流程的升级与完善过程中,知识挖掘技术扮演了不可或缺的角色。通过对档案管理全过程中产生的各类数据实施深入分析和挖掘,能揭示现有工作流程存在的效率低点和缺陷,为流程改进提供策略建议。比如,借助对档案借阅、归还、维护等步骤的数据洞察,可以识别出制约工作效率的核心要素,进而对流程架构进行合理化调整。

档案管理流程改进示例数据表1-1

环节	当前效率(%)	问题识别	改进措施	预期效果(效率提升 %)
档案分类	70	人工分类效率低,易出错	引入自动化分类系统	90
档案借阅	60	借阅流程繁琐,等待时间长	优化借阅流程,简化手续	85
档案归还	65	归还管理混乱,记录不全	实施电子归还记录系统	80
档案维护	75	维护频率不均,资源浪费	利用数据分析优化维护计划	88
信息检索	50	检索系统不智能,耗时长	应用智能检索和推荐系统	90

通过上述策略,档案管理流程的效能与品质将获得显著增强,进而促进档案信息的深层次开发与有效运用。

## 3 知识挖掘技术在档案信息深度开发中的实施步骤

### 3.1 数据预处理与清洗

数据预处理和清洗在知识挖掘中扮演着至关重要的角色,其首要任务是提升数据品质,确保挖掘出的信息既精确又可信。数据的源头至关重要,它源自多元化的档案体系和数据库,首先需实施有效的整合策略,将这些纷繁复杂的资料初步归一化。在这个过程中,去重是基础操作,通过剔除重复的记录,致力于维护数据的独特性,避免信息冗余带来的混淆。面对数据中的空缺,可能的处理策略包括摒弃整个含有缺失值的条目,或者采用统计手段如平均值或中位数来填充,或者借助高级的预测模型来弥补空白。数据转换则如同艺术中的调色板,通过规范化和标准化,统一了数据的格式和度量标准,消除了因数据来源多样性导致的理解偏差。<sup>[4]</sup>经过这一系列精心的预处理和清洗步骤,数据焕然一新,从而为深入的知识挖掘提供了稳固且高质量的基石。

### 3.2 知识挖掘模型的构建

构建知识挖掘模型的过程是知识挖掘技术的核心策略,其精髓在于精准匹配档案数据的特性和挖掘需求,以选用最适宜的分析方法。这一范畴涵盖了多元化的模型设计,诸如用于精确分类的策略模型(如基于决策路径的决策树,集成学习的随机森林,以及支持向量机的精准划分),旨在揭示内在群体结构的群组模型(如K-means的标准化分群,以及层次聚类的递进组织),

以及揭示数据间关联和模式的规则模型(如Apriori的频繁项集挖掘,和FP-Growth的高效增长算法)。回归模型则在预测数值趋势和关联性方面发挥关键作用。为了提升模型性能,特征选择和特征工程起着至关重要的角色,它们通过精挑细选核心特征并创新性地构造新特征,从而提高模型的预测精度和执行效率。后续的模型优化阶段,通过交叉验证和严格的模型评估,确保了模型在实际应用中的稳定性和效能。

知识挖掘模型构建步骤表2-1

步骤	内容	方法
算法选择	根据数据特征和目标选择合适算法	决策树、K-means、Apriori等
特征选择	筛选对模型有重要影响的特征	特征选择算法
特征工程	生成新的特征,提升模型表现	特征转换、组合
模型评估	通过交叉验证和评估调整模型	交叉验证、模型调优

通过以上细致而系统的步骤,得以创建出在实际中表现出高效、精准和可靠特性的知识挖掘模型,为数据分析提供了强有力的工具。

### 3.3 档案信息深度挖掘

模型构建完毕后,紧接着是档案信息的深层次探索阶段。该阶段涵盖了数据的实际操作、模型实施以及智慧提炼。将经过预处理的数据导入建立的挖掘模型,进行全面分析,以揭示数据内部隐藏的价值线索和结构。这个过程涉及多个步骤,如分门别类、群组划分、关联规律探索以及趋势预估。比如,借助分类算法,可实现档案的自动化分类;运用聚类技术,能揭示档案间的内在联系;利用关联规则,能够挖掘档案中的常见组合和规律;而通过回归分析,则能预测档案数据的未来走向。这些模型的运用,使得档案信息中的关键模式和法则得以显性化,为档案管理和使用提供了坚实的理论支撑和决策指导。此外,深度挖掘还能识别档案管理中的异常现象和潜在问题,从而增强管理的覆盖范围和精确度。<sup>[5]</sup>

### 3.4 知识表示与可视化

知识表示与可视化构成了知识挖掘流程的最终阶段,它通过艺术化的手法揭示挖掘的精华,使信息的应用与理解更为便捷。知识的呈现方式多种多样,涵盖了文本叙述、视觉图表和复

杂的知识网络等形式,力求以最简洁且直观的方式揭示隐藏的复杂性。诸如Tableau、Power BI和Gephi这样的可视化工具,如同魔术师的手法,能将数据和智慧转化为迷人的视觉盛宴。例如,知识图谱作为其中一种手段,能够精细地揭示档案之间的连接网络以及主题的分布格局,从而帮助使用者迅速洞悉数据的内在架构和规律。而柱状图、饼图和热力图等传统图表,则以其直观的方式揭示分类结果的分布、频繁事件的序列和趋势的演变,极大地增强了人们对数据直觉的理解。动态交互式的可视化设计,赋予用户参与探索的可能,使得知识的表达更为生动且易懂,为档案信息的深度解读和创新应用提供了强大的视觉辅助工具。

## 4 结语

在档案信息的深层次开发中,知识挖掘技术的运用开创了全新的可能性。涉及数据前期的准备和净化、知识挖掘架构的设计、档案信息的深入探析以及知识的表达与直观展示,这些步骤均能显著增强档案管理的效能和精确度,从而有力地辅助决策过程。展望未来,随着数据技术的进步和档案数据的不断增长,知识挖掘技术的角色将日益凸显。不过,应用此技术的同时,必须面对并解决数据隐私和安全性等关键问题,以保证操作的合法性。知识挖掘在档案信息深挖领域的应用潜力巨大,有待更深入的研究和实验,以促进档案管理向现代化和智能化转型。

## 参考文献

- [1]赵春全,李宗才.知识挖掘技术在高校档案管理中的应用研究[J].中国战略新兴产业,2022(20):191-193.
- [2]刘禹辰,程奕心,许志伟,等.网上数字档案大数据分析中的知识挖掘技术[J].电子技术与软件工程,2020(5):2.
- [3]韩蔚.知识挖掘技术在企业内部审计中运用研究[J].全国流通经济,2021(26):3.
- [4]刘岚萍,旷文俊.档案数据深度挖掘与精细化管理:大数据时代档案信息化建设策略[J].知识经济,2020(2):2.
- [5]王宪东,孙洪鲁,赵琳.基于智能词表的电子档案和政务数据深度挖掘技术及应用工具[J].中国档案,2023(12):56-57.

## 作者简介:

曹虹(1982-),女,汉族,湖南长沙人,硕士,副研究馆员,研究方向:档案管理与档案学研究。