

# Research on the construction and application of archival value appraisal model based on big data analysis

Quangang Pan

Hulle Town People's Government, Zalait Banner, Xingan League, Inner Mongolia, 137617, China

## Abstract

With the advent of the information age, archive management is increasingly challenged by the need to store and manage vast amounts of data. Archives, which are crucial for history, culture, law, and administration, play a vital role in their value assessment, which is essential for effective archive management. However, traditional methods of assessing archive value are highly subjective and lack systematic quantitative standards, making it difficult to meet the growing volume of archive information and the complex, ever-changing needs of value assessment. The aim of constructing an archive value assessment model based on big data analysis is to introduce big data technology and integrate the multi-dimensional characteristics of archives to establish a more scientific and objective system for evaluating archive value. This paper analyzes the application of big data technology in archive value assessment, explores the methods and steps involved in building such a model, designs a corresponding framework, and conducts applied research based on real-world cases.

## Keywords

big data analysis, archival value, appraisal model, archival management, data mining

# 基于大数据分析的档案价值鉴定模型构建与应用研究

潘全纲

内蒙古自治区兴安盟扎赉特旗胡尔勒镇人民政府, 中国·内蒙古 兴安盟 137617

## 摘要

随着信息时代的到来, 档案管理逐渐面临着大量数据存储和管理的挑战。档案作为历史、文化、法律、行政等领域的重要资料, 其价值的鉴定对档案管理工作至关重要。然而, 传统的档案价值鉴定方式存在较大的主观性, 且缺乏系统的量化标准, 难以适应日益增长的档案信息量和复杂多变的鉴定需求。基于大数据分析的档案价值鉴定模型的构建, 旨在通过引入大数据技术, 结合档案的多维度特征, 建立更加科学、客观的档案价值评价体系。本文通过分析大数据技术在档案价值鉴定中的应用, 探讨了大数据分析在构建档案价值鉴定模型中的方法与步骤, 设计了相应的模型框架, 并基于实际案例进行了应用研究。

## 关键词

大数据分析, 档案价值, 鉴定模型, 档案管理, 数据挖掘

## 1 引言

档案管理作为信息管理的重要组成部分, 在现代社会的各个领域发挥着至关重要的作用。档案不仅记录了历史事件, 反映了社会变迁, 还在政府决策、社会治理、企业发展等方面具有重要的参考价值。然而, 随着社会的不断进步和信息技术的飞速发展, 档案的数量和种类急剧增加, 传统的档案管理模式和鉴定方法已经无法适应新的需求<sup>[1]</sup>。档案的价值鉴定一直是档案管理中的一个关键问题。合理的档案价值鉴定能够有效识别出具有历史、法律、文化等重要价值的档案, 进而对其进行科学地保存和利用。

【作者简介】潘全纲(1984-), 男, 蒙古族, 中国内蒙古兴安盟人, 本科, 中级馆员, 从事档案管理研究。

传统的档案价值鉴定通常依赖于专家的主观判断, 存在一定的误差和偏差, 而且缺乏量化的评价标准。近年来, 随着大数据技术的不断发展, 利用大数据分析手段来提高档案价值鉴定的科学性和准确性成为一种新的趋势。大数据分析可以处理海量数据, 发现数据之间的潜在关系, 并通过数据挖掘和模型构建, 提供客观、量化的评价依据<sup>[2]</sup>。因此, 构建基于大数据分析的档案价值鉴定模型, 具有重要的理论意义和实践价值。

## 2 档案价值鉴定的传统方法及其不足

档案价值鉴定作为档案管理的一项核心工作, 其主要目的是根据档案的使用价值、历史价值、文化价值等因素, 对档案的保存、利用及处置进行科学决策。传统的档案价值鉴定方法主要依赖于专家经验判断, 通过对档案内容、形式、

背景等因素的分析,来评估其价值。然而,这种方法存在一定的局限性,首先,由于专家的主观因素较多,导致不同专家对同一份档案的价值判断可能存在较大差异。其次,传统方法对于档案价值的衡量标准较为模糊,缺乏统一和量化的评价标准,导致鉴定结果的不稳定性<sup>[3]</sup>。此外,随着档案数量的激增,人工鉴定的工作量大大增加,造成档案价值鉴定工作效率低下,且容易出现遗漏和偏差。

因此,如何克服传统档案价值鉴定方法中的不足,利用现代信息技术提高档案价值鉴定的科学性、准确性和效率,成为当前档案管理领域的研究热点。

## 2.1 专家主观判断的局限性

在传统的档案价值鉴定过程中,专家的主观判断一直占据着核心地位。专家通过综合档案的内容、历史背景、使用价值等多方面因素,给出档案的价值等级或评估结果。然而,这一过程存在着明显的局限性。专家的判断通常受到其个人经验、知识储备、认知视角、判断思维方式等因素的影响,而不同专家在面对相同档案时,往往得出不同的评定结论。这种高度依赖个人主观因素的评估方式,缺乏统一的标准和客观依据,导致档案的价值判断不够科学和精准。专家之间的评估差异使得档案价值鉴定的结果缺乏一致性,难以形成标准化、可操作性强的评价体系。此外,专家的工作负担较重且对其判断结果的精确度要求较高,在面对大量档案时,主观判断更容易受到心理疲劳、判断偏差等因素的影响,进而影响评估结果的可靠性。

## 2.2 缺乏量化评价标准

传统的档案价值鉴定大多依赖定性的分析方法,缺少量化的评价标准,这也是其局限性的一个重要方面。尽管一些国家和地区已经制定了档案价值鉴定的指导标准和方法,但这些标准仍停留在理论或初步阶段,缺乏具体的操作性和量化的评估指标。在实际操作中,许多档案的价值评定缺乏科学、系统地量化分析,导致评估结果的主观性较大<sup>[4]</sup>。例如,档案的历史背景、文化价值等因素,往往没有具体的量化标准,专家需要根据其对历史文化的理解和经验做出判断,而这在处理不同类型档案时可能造成差异较大的评价结果。而档案的种类和内容往往复杂多样,传统的评估方法无法全面、精准地反映档案的实际价值,尤其是在处理大量档案数据时,传统方法往往无法全面考虑档案的各种属性,容易遗漏重要信息和潜在价值。因此,如何制定出统一且具操作性的量化评价标准,成为当前档案管理领域亟待解决的问题。

## 2.3 人工鉴定的工作量大且效率低下

随着信息化时代的到来,档案的数量和种类迅速增加。传统的人工鉴定方式面临着巨大的工作量压力,特别是在需要评估大量档案的场合,人工鉴定不仅耗时且容易出现遗漏和错误。在档案管理中,由于档案种类繁多、内容复杂,人工鉴定所需的时间和精力成倍增长,尤其在档案数量巨大的

情况下,人工鉴定不仅造成管理成本的增加,还容易导致评估结果的错误或不完整,进而影响档案管理工作的质量。此外,人工鉴定的低效率和高误差率使得档案管理工作面临着严峻的挑战。在传统的人工鉴定方法中,鉴定人员常常需要逐一每份档案进行分析和判断,耗费大量时间和精力,且难以保证每一份档案都能被全面、精准地评估。随着档案数量的不断增加和信息技术的飞速发展,传统的人工鉴定方式已经逐渐无法满足现代档案管理的需求,迫切需要引入新技术、新方法来提高档案价值鉴定的效率和准确性。现如今,人工鉴定已经不再是唯一的解决办法,通过自动化、智能化的技术手段能够有效提升档案鉴定的工作效率,同时也能降低人为错误的发生。

## 3 基于大数据分析的档案价值鉴定模型构建

基于大数据分析的档案价值鉴定模型,通过引入数据挖掘、机器学习、人工智能等现代技术手段,利用海量数据对档案的价值进行科学的评估。大数据分析能够挖掘档案数据之间的潜在关系,通过对档案内容、形式、背景等多维度信息的全面分析,为档案的价值评定提供客观、量化的依据<sup>[5]</sup>。大数据分析不仅能够克服传统档案价值鉴定方法中的主观性问题,还能显著提高鉴定效率,提升档案管理的科学性和精准性。

### 3.1 数据收集与处理

构建基于大数据分析的档案价值鉴定模型的第一步是数据收集和处理。这一过程的核心任务是确保收集到的档案数据具有高质量和多样性。为了全面评估档案的价值,需要从多个角度进行数据收集,主要包括档案的基本信息、档案的内容、历史背景、使用情况、存储情况等。这些数据能够为后续的分析提供全面的支持。档案的基本信息包括档案的名称、类型、创建时间、存储方式等,这些有助于了解档案的物理属性。档案的内容则需要分析档案所含有的知识价值、技术价值等,历史背景可以提供档案的时代价值、文化背景等,而使用情况则反映了档案的实际应用价值。此外,信息的来源多样性要求在不同来源间协调采集,可能涉及政府档案库、企业文件、图书馆存档等。

数据处理是确保分析结果可靠性的关键步骤。在这一阶段,首先需要对收集到的数据进行清洗和去重,去除冗余信息,确保数据的一致性与质量。数据归一化操作有助于将不同尺度的变量转换为统一标准,保证数据在后续分析中的公正性和科学性。然后,根据分析目标对数据进行分类、标签化,使得数据能够按属性有序排列,方便后续建模使用。

### 3.2 特征提取与模型构建

在数据收集和处理的基础上,下一步是进行特征提取。特征提取是大数据分析中至关重要的一步,关系到模型的准确性和分析效果。特征提取的目标是从大规模的数据中提炼出最具价值的元素,这些特征直接影响到档案价值评估的

精度。通过深度挖掘档案数据，可以提取出与档案价值相关的特征指标。例如，档案的历史背景是影响档案价值的一个重要因素，历史上某些档案由于其特定的历史事件或文化背景，具有更高的价值；文献引用量是衡量档案学术价值的一项指标，频繁引用的档案通常有较高的研究价值；使用频率是衡量档案在实际应用中的价值，档案的频繁使用往往表明其在实践中的重要性；存储时间长短、存储环境等也是影响档案保存完整性及其价值的因素。

在特征提取完成后，下一步是选择适合的算法来构建档案价值鉴定的数学模型。常见的模型构建方法包括决策树、支持向量机（SVM）、神经网络等。这些方法通过学习档案的特征信息，建立档案价值的评估体系。决策树通过对特征进行二叉划分，建立简单的分类模型；支持向量机则利用高维空间中的边界最优化进行分类，对大规模数据具有较强的泛化能力；神经网络通过模拟人脑的神经元连接，进行深度学习，能够处理更复杂的模式识别任务。通过对这些算法的综合使用，可以根据档案的特征信息，评估其价值。

### 3.3 模型训练与优化

在完成模型构建之后，必须进行模型训练和优化，以确保其精确性和高效性。模型训练是通过历史数据对构建的模型进行拟合，从而让模型能够预测未见数据的准确结果。通过优化算法参数，模型可以在处理不同类型的档案时提高评估的准确性。为了提高模型的稳定性和准确性，训练过程可能采用交叉验证技术，确保模型在不同数据集上都能保持良好的表现。

在模型训练的过程中，还可能采用集成学习和深度学习等先进技术。集成学习通过组合多个基模型来提高预测精度，而深度学习则能够通过多层神经网络从更高层次提取复杂的特征，进行精确的分类和预测。通过这些技术的结合，模型能够不断提升其性能，适应不同的档案数据和评估需求。

此外，优化工作也包括防止过拟合和欠拟合的问题。过拟合会导致模型在训练数据上表现过于完美，但无法有效处理新数据；而欠拟合则意味着模型未能充分捕捉到数据中的规律，导致预测能力弱。通过定期调整算法和模型参数，保持模型的灵活性，能够有效避免这些问题的发生。

## 4 基于大数据分析的档案价值鉴定模型的应用研究

通过基于大数据分析的档案价值鉴定模型，能够有效

解决传统方法中的问题，提高档案价值鉴定的效率和准确性。在实际应用中，模型能够根据不同类型的档案数据，快速生成档案的价值评估结果。例如，在一些档案管理单位，采用该模型可以迅速对大量的档案进行价值评定，帮助管理人员合理规划档案的保存和利用。此外，该模型还能够根据档案的使用情况和历史数据，不断进行自我优化和调整，提高档案鉴定的精确度和适应性。

## 5 基于大数据分析的档案价值鉴定模型的优化策略

尽管基于大数据分析的档案价值鉴定模型在实践中取得了较好的效果，但仍然面临一些挑战和优化空间。首先，数据质量的提高是优化模型的关键。由于档案数据的来源广泛，数据质量参差不齐，因此需要不断完善数据采集和处理的标准，确保数据的准确性和一致性。其次，模型的可解释性也是一个需要关注的问题。当前一些大数据分析模型，尤其是深度学习模型，虽然在预测精度上具有优势，但其内部的计算过程较为复杂，缺乏透明度。因此，未来可以通过引入可解释性 AI 技术，提高模型的可解释性，帮助使用者更好地理解模型的评估结果。

## 6 结语

基于大数据分析的档案价值鉴定模型为档案管理提供了新的思路和方法，通过引入数据挖掘和机器学习技术，能够有效提高档案价值鉴定的科学性和准确性。然而，在实际应用过程中，仍然面临数据质量、模型可解释性等方面的挑战。未来，随着技术的不断进步，基于大数据的档案价值鉴定模型将更加完善，为档案管理提供更为有效的支持，推动档案管理工作的科学化和现代化。

### 参考文献

- [1] 陶永昌.浅析房屋交易与产权档案管理工作的创新策略[J].黑龙江档案,2024,(06):130-132.
- [2] 王艳红.浅谈计划生育档案管理工作的提升路径[J].黑龙江档案,2024,(06):100-102.
- [3] 李林栋.智慧医院建设背景下医院档案管理的探索与实践[J].黑龙江档案,2024,(06):109-111.
- [4] 杨会.新时代街道干部档案管理工作研究[J].黑龙江档案,2024,(06):172-174.
- [5] 张华丽.关于提升事业单位干部人事档案管理效率的几点思考[J].黑龙江档案,2024,(06):178-180.