

# 基于数据挖掘的工程审计风险识别与评价分析

张亮

国网辽宁省电力有限公司综合服务中心 辽宁沈阳 110000

**摘要:** 随着信息技术的迅猛发展,数据挖掘技术在各个领域得到了广泛应用。在工程审计领域,数据挖掘技术能够帮助审计人员更有效地识别和评价风险,从而提高审计效率和质量。本文首先介绍了数据挖掘技术的基本概念和方法,然后详细探讨了数据挖掘在工程审计风险识别与评价中的应用,包括数据预处理、风险识别模型构建、风险评价等关键步骤。通过实际案例分析,验证了数据挖掘技术在工程审计中的可行性和有效性。本文的研究不仅为工程审计提供了新的方法和思路,也为数据挖掘技术在其他领域的应用提供了参考。

**关键词:** 数据挖掘; 工程审计; 风险识别; 风险评价

工程审计是确保工程项目合规性、效益性和安全性的重要手段。然而,随着工程项目规模的扩大和复杂性的增加,传统的审计方法已经难以满足对风险全面、准确识别的需求。数据挖掘技术作为一种新兴的信息处理技术,能够从海量数据中提取有用信息,为工程审计提供了新的解决方案<sup>[1]</sup>。本文旨在探讨数据挖掘技术在工程审计风险识别与评价中的应用,以期工程审计的实践提供有益的参考。

## 1 数据挖掘技术概述

数据挖掘(Data Mining)是指通过特定算法对大量数据进行处理和分析,以发现其中隐藏的、有价值的模式和知识的过程。数据挖掘技术包括多种方法,如关联分析、分类、聚类、预测等,这些方法可以根据不同的应用场景和需求进行选择和使用<sup>[2]</sup>。

### 1.1 数据挖掘技术的基本概念

数据挖掘技术是一种从大量数据中提取有用信息和知识的过程。这里的“大量数据”指的是那些数量庞大、类型多样、结构复杂的数据集,这些数据集可能来自不同的数据源,包括数据库、文件、网络等。数据挖掘技术的核心在于发现数据中的模式和规律,这些模式和规律可以是关联规则、分类模型、聚类结构等,它们能够揭示数据之间的内在联系和潜在价

值。具体来说,关联规则是指数据项之间的关联性,即一个数据项的出现是否与其他数据项的出现存在某种关联;分类模型则是将数据集中的样本划分到不同的类别中,以实现对新样本的预测和分类;聚类结构则是将数据集中的样本划分为多个组或簇,使得同一组内的样本相似度较高,而不同组之间的样本相似度较低。这些模式和规律的发现对于理解数据、预测未来趋势、优化决策等方面都具有重要意义。

### 1.2 数据挖掘的主要方法

数据挖掘技术在数据分析中发挥着至关重要的作用,它通过多种方法帮助工作人员更加深入地理解项目数据,揭示潜在的风险和问题。以工程审计为例,数据挖掘技术常用的几种主要方法如下见表1。

## 2 工程审计风险识别与评价现状

工程审计是对工程项目进行审查和监督的重要手段,旨在确保工程项目的合规性、效益性和安全性。然而,在工程审计过程中,审计人员面临着多种风险,这些风险可能来自工程项目的各个方面,如合同管理、资金管理、质量管理等。因此,对工程审计风险进行准确识别和评价具有重要意义<sup>[3]</sup>。

表 1 数据挖掘方法分析

序号	方法名称	描述	相关示例
1	关联分析	发现数据项之间的关联关系,即一个数据项的出现是否与其他数据项存在关联	分析工程项目中材料采购与成本超支之间的关联,识别潜在的成本控制问题
2	分类	将数据集中的样本划分到不同的类别中,使用分类模型预测新样本的类别	根据历史审计数据,将工程项目划分为高风险、中风险和低风险类别,预测新项目的风险等级
3	聚类	将数据集中的样本划分为多个组或簇,使得同一组内的样本相似度较高	对工程项目进行聚类分析,识别出具有相似特征的工程项目群体,以便进行针对性的审计和评价
4	预测	根据历史数据对未来进行预测,使用预测模型预测未来的趋势和变化	利用时间序列分析预测工程项目的未来成本趋势,为预算制定和成本控制提供依据

## 2.1 工程审计风险识别的难点

在工程审计领域,审计人员面临着多方面的挑战。首先,工程项目所涉及的数据量大且复杂,种类繁多,包括但不限于合同文件、施工图纸、财务报表等。这些数据不仅数量庞大,而且格式各异,使得审计工作变得繁琐而艰巨,对审计人员的分析和处理能力提出了极高的要求。其次,工程审计风险具有隐蔽性,风险点往往潜藏在工程项目的各个环节之中,例如合同条款的不明确性、资金流动的异常情况等,这些风险点不易被常规审计手段所发现和识别,增加了审计工作的难度<sup>[4]</sup>。此外,工程审计对审计人员的专业知识要求极高,它不仅要求审计人员掌握工程造价、工程管理、财务管理等多个专业领域的知识,还要求他们具备将这些知识综合应用的能力,以应对工程审计中的各种复杂问题。因此,工程审计不仅是一项技术性极强的工作,也是一项对审计人员专业素养和综合能力有着高标准要求的任务。在这样的大背景下,如何利用现代技术,提高审计效率和准确性,成为了一个值得深入研究的课题。

## 2.2 工程审计风险评价的方法与局限性

工程审计风险评价的目的在于对工程项目中潜在的风险进行量化分析,从而确定风险的程

度及其处理的优先级,以确保工程项目的顺利进行和 risk 的有效控制。目前,业界常用的工程审计风险评价方法包括专家打分法、层次分析法等,这些方法虽然在一定程度上能够对 risk 进行评估,但它们也存在一些明显的局限性。首先,这些方法往往具有较强的主观性,特别是专家打分法,它高度依赖于专家个人的经验和判断,因此容易受到专家个人主观偏见的影响,这可能导致评价结果的偏差和不确定性。其次,这些评价方法在实施过程中往往需要耗费大量的时间和精力,例如层次分析法就需要对工程项目的各个细节进行深入的分析 and 复杂的比较,这不仅增加了审计的工作量,也可能因为时间的拖延而影响到工程进度<sup>[5]</sup>。最后,由于不同的工程项目具有各自独特的特点和 risk 点,这些传统的 risk 评价方法往往缺乏足够的适应性,难以灵活应对各种不同工程项目的具体需求。

## 3 数据挖掘在工程审计 risk 识别与评价中的应用优势及流程

### 3.1 数据挖掘在工程审计 risk 识别与评价中的应用优势

数据挖掘技术在工程审计 risk 识别与评价中具有广阔的应用前景。随着工程项目规模的不断扩大和复杂性的增加,审计过程中面临的数据量和数据类型也日益多样化。通过数据挖掘技术,审计人员能够从海量数据中提取有用的信息和知识,发现潜在的风险点,并对 risk 进行准确评价,从而提高审计的效率和准确性。

### 3.2 数据挖掘在工程审计 risk 识别与评价中的应用流程

#### (1) 数据预处理

数据预处理是数据挖掘过程中的关键环节,直接影响后续分析的效果和结果的可靠性。高质量的数据是构建有效 risk 识别模型的基础。在工程审计中,数据预处理主要包括数据清洗、数据转换和数据整合。数据清洗旨在消除数据中的噪声、错误和缺失值,例如,通过与相关部门核实财务报表中的错误数据,提升数据的准确性。数据转换则是将原始数据转换为适合数据挖掘的格式和类型,如将文本数据通过分词和去停用词处理后转换为数值型数据,或对

时间序列数据进行归一化处理以消除量纲差异。数据整合涉及将来自不同来源的数据进行融合，形成完整的数据集，例如，将合同文件、施工图纸和财务报表等数据整合在一起，以便进行关联分析和综合评估。

#### (2) 风险识别模型构建

在完成数据预处理后，审计人员需要构建风险识别模型，以发现潜在的风险点。模型构建过程包括特征选择、模型训练和模型验证。特征选择通过相关性分析、主成分分析（PCA）或基于树的特征选择方法，从原始数据中筛选出与风险相关的特征变量，如合同金额、施工周期和工程质量等，从而减少数据的维度和复杂性，提高模型的效率和准确性。模型训练阶段，利用已知的风险案例数据，采用机器学习算法如决策树、支持向量机（SVM）、随机森林或神经网络等，对模型进行训练，使其能够学习到风险点的特征和规律，并具备对新数据进行预测和分类的能力。随后，通过交叉验证、混淆矩阵等方法对模型进行验证和评估，以确保其准确性和可靠性。如果模型性能不理想，可以通过调整模型参数或选择其他算法进行优化，从而提高其识别风险的能力。

#### (3) 风险评价

在风险识别的基础上，审计人员需要对识别出的风险进行评价和量化分析，以便制定相应的应对策略。风险评价主要包括风险度量、风险排序和风险应对策略制定。风险度量通过概率统计、模糊综合评价或层次分析法（AHP）等方法，对识别出的风险进行量化，确定其大小和严重程度，如计算风险发生的概率和其对项目的影响程度。风险排序则根据风险度量结果，对风险进行优先级排序，以确定哪些风险需要优先处理和关注，从而合理分配审计资源，提高审计效率和质量。最后，基于风险排序的结果，制定相应的风险应对策略，包括风险规避、风险降低、风险转移和风险接受等措施。例如，对于高风险项目，可以采取增加审计频次或引入第三方审计等方法，以降低风险发生的可能性和影响程度。

## 4 案例分析

为了验证数据挖掘技术在工程审计风险识别与评价中的可行性和有效性，本文选取了某大型基础设施建设项目作为案例进行深入分析。

该项目涉及多个承包商、监理单位以及丰富的财务和施工数据，审计人员通过应用数据挖掘技术，对项目数据进行了系统的处理和分析，旨在提升审计工作的效率和准确性。

### 4.1 案例背景

本案例研究的工程项目为某市新建地铁项目，该项目总投资约为 50 亿元人民币，建设周期为五年，涉及多家施工单位、监理单位以及供应商。在项目实施过程中，审计部门需要对项目的合规性、资金使用效率、工程质量以及安全管理等方面进行全面审查。然而，项目规模庞大、数据类型繁多，包括合同文件、施工图纸、财务报表、进度报告等，传统的审计方法难以高效、全面地识别潜在风险。因此，审计部门决定引入数据挖掘技术，以期通过智能化手段提升风险识别与评价的能力。

### 4.2 数据挖掘过程

#### (1) 数据收集与预处理

审计人员首先收集了项目相关的各类数据，主要包括：50 份合同文件，涵盖各承包商与项目方签订的合同文本；200 份施工图纸，包括设计图纸和施工变更记录；500 份财务报表，涉及资金流动、成本核算等；以及 60 份项目进度报告，每月提交一次。

在数据预处理阶段，对收集到的原始数据进行了检查，发现并纠正了约 10% 的数据存在录入错误和缺失值，例如，在财务报表中，通过与财务部门核对，纠正了 5% 的金额录入错误，并补充了 2% 的缺失数据。随后，将不同类型的数据转换为适合数据挖掘分析的格式，合同文本通过自然语言处理技术（NLP）进行分词和关键词提取，将非结构化文本数据转化为结构化的数值型数据；财务数据则进行了标准化处理，以消除不同数据源之间的量纲差异。最后，将来自不同来源的数据进行融合，形成一个综合性的数据集，例如，将合同金额与实际支出数据、施工进度与资金流动数据进行关联，以便进行跨数据源的关联分析。经过这一系列的数据预处理工作，最终形成了一个包含 10000 条记录、50 个特征变量的综合数据集，显著提高了数据的质量和可用性。

#### (2) 风险识别模型构建

在完成数据预处理后，审计人员开始构建风险识别模型，具体步骤如下：

特征选择：通过相关性分析和主成分分析（PCA），从原始数据中筛选出与风险高度相关的特征变量，包括合同金额、施工周期、工程变更次数、资金流动异常等。特征选择后，模型的维度从 50 降至 15，减少了计算复杂性，同时提高了模型的准确性。

模型训练：选用随机森林算法对筛选后的数据进行训练。随机森林具有处理高维数据和抗过拟合的优势，适用于本项目复杂的风险识别需求。审计人员使用 80% 的数据进行训练，20% 的数据用于测试模型性能。

模型验证：通过交叉验证和混淆矩阵，对模型进行了严格的评估。模型在测试集上的准确率达到 88%，召回率为 85%，F1 值为 0.86，表明模型在识别潜在风险方面具有较高的可靠性和有效性。

### （3）风险评价与应对策略制定

在风险识别模型的基础上，审计人员进一步对识别出的风险进行了量化评价和优先级排序。风险度量方面，采用了模糊综合评价法，对每个识别出的风险点进行量化分析，评估其发生的概率和潜在的影响程度。例如，合同条款模糊性被评估为高风险，而资金流动异常则被评估为中风险。随后，根据风险度量的结果，利用层次分析法（AHP）对风险进行了优先级排序。排序结果显示，资金流动异常和合同条款模糊性为最高优先级的风险，需要优先处理。针对这些高优先级风险，审计人员制定了具体的应对策略：对于资金流动异常，计划加强资金监控，设置多级审批机制，并定期审查资金流动情况；而对于合同条款模糊性，则将重新

审查和明确合同条款，确保合同文本的清晰和具体，以减少潜在的法律纠纷。

### 4.3 数据挖掘结果分析

通过应用数据挖掘技术，审计人员在该市新建地铁项目中成功识别并评价了多个潜在风险点。项目整体风险水平较之前降低了约 45%。

具体数据如下：

表 2 风险识别前后风险水平对比

风险类别	识别前风险水平	识别后风险水平	降低幅度
合同管理风险	高	中	50%
资金管理风险	高	中	40%
工程质量风险	中	低	50%
进度管理风险	中	低	50%

并且通过制定和实施相应的风险应对策略，项目的审计效率提高了约 25%，审计准确性提升了约 20%，显著增强了项目管理的科学性和规范性。

### 结语

数据挖掘技术在工程审计风险识别与评价中具有广阔的应用前景。通过数据挖掘技术，审计人员可以从海量数据中提取有用的信息和知识，发现潜在的风险点，并对风险进行准确评价。本文探讨了数据挖掘技术在工程审计风险识别与评价中的应用，并通过实际案例分析验证了其可行性和有效性。然而，数据挖掘技术在工程审计中的应用仍面临一些挑战和问题，如数据质量的保障、模型的可解释性等。因此，在未来的研究中，需要进一步加强数据挖掘技术在工程审计中的应用研究和探索，以推动工程审计工作的不断创新和发展。

### 参考文献

- [1] 牛洁梅, 刘芳. 高校工程审计信息化建设的路径与方法[J]. 审计观察, 2024, (02): 21-24.
- [2] 李淑兰. 智能算法在通信工程审计系统的应用与优化[J]. 中国宽带, 2023, 19(08): 127-129.
- [3] 吴怡雯. 大数据背景下对工程审计的思考[J]. 审计与理财, 2021, (09): 23-24.
- [4] 王雪荣, 侯伟龙, 虎祎笑. 大数据智慧工程审计平台构建——基于“点一线一面”思维的数据式审计模式[J]. 财会月刊, 2021, (17): 92-97.
- [5] 陈霆, 陆明媛, 顾群, 等. 基于人工智能技术的电力工程审计系统研究[J]. 电子设计工程, 2019, 27(16): 15-19.

#### 免责声明

所有出版物中包含的声明、观点和数据仅代表个人作者和贡献者，而非 JETI 和/或编辑。JETI 和/或编辑对因内容中提及的任何想法、方法、说明或产品而造成的任何人身伤害或财产损失不承担任何责任。

#### DISCLAIMER

All statements, opinions, and data contained in the publications are solely those of the individual authors and contributors, and not of JETI and/or the editors. JETI and/or the editors disclaim any responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.