

基于决策树的电力工程造价异常数据挖掘技术

张欣然

内蒙古送变电有限责任公司 内蒙古 呼和浩特 010020

摘要: 在电力工程造价管理中,异常数据影响决策准确性。本文基于决策树技术,分析与电力工程造价数据的适配逻辑,阐述造价异常数据类型、形成机制及挖掘前提。详细介绍基于决策树的造价异常数据挖掘技术路径,包括数据预处理、模型构建与参数设置、异常识别流程。最后提出挖掘技术的优化方向,涵盖决策树模型、数据预处理技术优化及挖掘技术与造价管理的融合,为电力工程造价管理提供可靠数据支撑。

关键词: 电力工程造价;决策树技术;异常数据挖掘;技术路径;优化方向

引言:电力工程造价管理涉及项目建设全周期费用,数据多源异构且易受外界干扰,存在潜在异常数据隐患。异常数据若未及时识别与处理,会导致造价分析偏差,影响项目决策与成本控制。决策树技术依托树状拓扑结构,能实现数据分类与识别,具有可解释性强、无需复杂数据预处理等优势,与电力工程造价异常数据挖掘需求天然适配。深入研究基于决策树的电力工程造价异常数据挖掘技术,对提升造价管理准确性、保障项目顺利实施具有重要意义。

1 电力工程造价数据与决策树技术的适配逻辑

1.1 电力工程造价数据的构成与运行特征

电力工程造价数据涵盖项目建设全周期内各类费用相关数据,核心构成包括静态投资与动态投资对应的各类数据维度,静态投资相关数据包含建筑工程费、设备购置费、安装工程费及其他相关费用数据,动态投资相关数据则涵盖建设期利息等波动类数据^[1]。这类数据来源呈现多源异构特征,既包括电力工程造价与定额管理总站发布的标准性数据,也包含工程建设各环节产生的实操性数据。运行特征上,数据兼具连续性与离散性,部分数据如设备购置费、建筑工程费呈现连续波动态势,受市场材料价格、工程规模等因素影响显著;同时存在明确的分层特征,不同电压等级、工程类型对应的数据分布存在明显差异,且易受外界因素干扰出现数据偏差,形成潜在的异常数据隐患,这一特征也决定了数据挖掘技术需具备较强的分层识别能力。

1.2 决策树数据挖掘技术的运行原理

决策树数据挖掘技术依托树状拓扑结构实现数据分类与识别,核心是通过递归分割将复杂的非线性决策过程拆解为一系列可解释的逻辑判断。技术运行核心在于节点分裂与纯度优化,根节点承载全量待分析数据,内部节点承担特征判断功能,根据数据特征的不同取值衍

生分支,引导数据进入下一级判别环节,叶子节点作为决策终点输出最终分类结果。节点分裂过程依赖特定量化指标,常用指标包括信息增益、信息增益率与基尼系数,通过这类指标量化特征对降低数据不确定性的贡献度,筛选最优分裂特征与分裂阈值。剪枝技术是技术运行中的关键环节,分为预剪枝与后剪枝两种方式,预剪枝通过设定停止条件限制树体深度,后剪枝则在完整树构建后合并冗余子树,以此平衡模型复杂度与泛化能力,避免过度捕捉数据噪声导致的识别偏差。

1.3 决策树技术与造价异常数据挖掘的适配性

决策树技术的结构特性与电力工程造价异常数据挖掘需求形成天然适配。造价数据的多维度分层特征与决策树的节点分裂逻辑高度契合,决策树可通过多级节点划分,精准匹配造价数据中不同费用类型、不同工程阶段的分层特征,实现对异常数据的分层识别。决策树无需复杂数据预处理的特性,能够适配造价数据多源异构、部分数据存在轻微缺失的实际情况,降低数据预处理成本的同时保留数据原始特征。其可解释性优势能够清晰呈现异常数据的识别路径,契合工程造价领域对数据挖掘过程可追溯的需求,便于梳理异常数据产生的逻辑关联。此外,决策树对连续型与离散型数据的兼容处理能力,能够全面覆盖造价数据的两类核心形态,通过合理选择分裂准则与剪枝策略,可进一步提升异常数据识别的精准度,适配造价数据波动频繁、异常类型多样的挖掘需求。

2 电力工程造价异常数据的类型与挖掘前提

2.1 电力工程造价异常数据的主要类型

电力工程造价异常数据可依据数据偏离程度与产生场景划分为明确类型,各类类型均对应造价管理中的典型数据异常表现^[2]。数值偏离型异常是最为常见的类型,指数数据取值显著偏离同类型、同规模电力工程项目的常

规数据区间,超出电力工程定额标准与市场合理波动范围,常见于设备购置费、建筑工程费等核心费用数据。逻辑矛盾型异常源于数据间内在关联冲突,即不同环节、不同维度的造价数据无法形成逻辑闭环,例如工程量数据与对应费用数据不匹配、施工进度数据与造价支付数据脱节等。缺失型异常指造价全周期数据采集过程中,关键数据项未完成记录或记录不完整,涵盖工程变更费用、材料单价明细等影响造价核算的核心数据,这类异常易导致造价分析出现偏差。

2.2 造价异常数据的形成机制

造价异常数据的形成是多因素共同作用的结果,核心与数据采集、管理及外部环境变化密切相关。数据采集环节的操作不规范会直接引发异常,包括采集人员对造价指标理解偏差、数据录入失误,以及多源数据对接时的格式不兼容的问题,导致原始数据出现失真。造价管理流程的不完善会加剧异常数据产生,部分项目未建立完善的数据校验机制,对数据的合理性、逻辑性未进行严格审核,使得偏差数据得以留存。外部环境波动同样会诱发数据异常,市场材料价格、人工费用的突发性波动,以及工程设计变更、施工工艺调整等,均会导致造价数据偏离初始基准,形成异常数据。

2.3 异常数据挖掘的基础数据处理要求

异常数据挖掘需以规范的基础数据处理为前提,处理过程需遵循电力工程造价数据管理规范,确保数据质量满足挖掘需求。数据清洗是核心处理环节,需剔除数据集中的错误数据、重复数据,补充缺失的关键数据项,通过合理的插值方法修正偏差数据,确保数据的准确性与完整性。数据标准化需统一各类造价数据的计量口径与表述规范,将不同来源、不同格式的数据转换为统一标准,消除因口径差异导致的挖掘偏差。特征筛选需聚焦造价异常识别相关的核心数据维度,剔除与异常判断无关的冗余数据,保留能够反映造价波动、逻辑关联的关键特征,为后续决策树挖掘技术的应用奠定基础,提升挖掘效率与精准度。

3 基于决策树的造价异常数据挖掘技术路径

3.1 造价数据预处理技术路径

造价数据预处理作为异常数据挖掘的前置核心环节,技术路径设计需紧密贴合电力工程造价数据多源异构、易受外界干扰、数据类型复杂的固有特点,严格遵循电力工程数据管理相关规范,分阶段有序推进且注重各环节衔接^[1]。数据筛选是预处理的首要步骤,需聚焦造价异常识别的核心需求,精准剔除与异常判断无直接关联的冗余数据项,重点保留建筑工程费、设备购置费、安

装工程费等核心费用数据,以及工程量、材料单价、人工费用、建设期利息等关联辅助数据,减少无关数据对后续挖掘过程的干扰,提升挖掘效率。数据清洗是保障数据质量的关键环节,需针对数据采集过程中易出现的录入错误、格式混乱、数据重复等问题开展系统性修正,采用线性插值、均值填补等规范且贴合造价数据特性的方法补充缺失关键数据,彻底剔除重复记录与明显失真数据,确保预处理后的数据具备准确性与完整性。数据标准化处理需立足电力工程定额标准,统一各类造价数据的计量口径与单位标准,将不同来源、不同格式、不同统计维度的造价数据转换为统一规范的形式,有效消除因口径差异、格式不统一带来的挖掘偏差,为后续决策树模型构建与异常识别提供高质量、高一一致性的数据支撑,夯实整个挖掘过程的基础。

3.2 决策树模型的构建与参数设置

决策树模型的构建与参数设置需紧密结合电力工程造价数据分层特征与异常识别需求,遵循数据挖掘技术规范,兼顾模型准确性、稳定性与可解释性,避免模板化构建。模型构建以预处理后的标准化造价数据为基础,通过相关性分析、特征重要性评估等方法完成特征变量精细化筛选,筛选出能真实反映数据波动规律、体现内在逻辑关联的关键特征,排除无关与冗余特征对模型构建的干扰。模型结构采用分层构建思路,适配造价数据分层特性优化节点设置,根节点优先选取对异常识别贡献度最高的特征变量,内部节点根据特征取值合理区间科学划分分支,叶子节点明确对应异常与正常数据的分类结果。参数设置重点优化分裂准则与剪枝参数,分裂准则优先选用信息增益率,兼顾特征区分度与模型稳定性;剪枝参数结合造价数据规模与分布特征合理设定,通过预剪枝限制树体深度、后剪枝合并冗余子树,平衡模型复杂度与泛化能力,杜绝过度拟合或拟合不足问题。

3.3 异常数据识别与挖掘的实施流程

异常数据识别与挖掘的实施流程需紧密衔接数据预处理成果与决策树模型构建结果,构建闭环式技术路径,每一步流程均遵循电力工程造价数据挖掘相关技术规范,确保挖掘过程可追溯、识别结果具备可靠性。流程正式启动后,将预处理完成的标准化造价数据精准输入已优化构建的决策树模型,通过模型内部节点的逐层逻辑判断,依据特征变量的具体取值完成数据分类操作,初步区分正常数据与疑似异常数据,为后续精准识别奠定基础。针对初步识别出的疑似异常数据,需开展进一步的精细化校验,结合电力工程定额标准、市场价格合理波动范围,对比疑似异常数据取值与常规数据区间的偏离

程度,综合分析数据偏离的合理性,排除正常市场波动、工程类型差异等因素导致的误判,提升异常识别的精准度。校验完成后,明确异常数据的具体类型、分布特征与取值范围,系统梳理异常数据产生的关联因素,形成完整、规范的异常数据挖掘结果^[4]。整个实施流程注重各步骤的无缝衔接,强化流程的规范性与严谨性,适配电力工程造价异常数据挖掘的实际需求,为后续造价管理优化提供可靠的数据支撑。

4 造价异常数据挖掘技术的优化方向

4.1 决策树模型的优化路径

决策树模型优化需围绕造价数据特征与挖掘需求,结合数据挖掘技术发展趋势持续完善。优化重点聚焦分裂准则与剪枝策略,突破传统单一准则局限,结合造价数据分层特性融合多分裂指标,优化特征筛选精度,减少无关特征对模型识别效果的干扰。针对模型泛化能力不足问题,引入集成学习思想,将多个决策树模型进行组合优化,通过模型融合降低单一模型的识别偏差,提升对复杂异常数据的识别能力。同时优化节点分裂阈值的动态调整机制,结合造价数据波动规律,实现阈值的自适应优化,适配不同工程类型、不同费用维度的异常识别需求,提升模型适配性与识别精准度。

4.2 数据预处理技术的完善方向

数据预处理技术完善需立足造价数据多源异构、易失真的特点,遵循电力工程数据管理规范优化处理流程。针对数据缺失问题,优化填补方法,结合造价数据关联性采用基于相似工程数据的自适应填补策略,替代传统单一插值方法,提升缺失数据填补的合理性。强化数据清洗的智能化水平,引入数据异常预判算法,提前识别采集过程中的潜在失真数据,减少后期数据修正成本。优化数据标准化流程,结合电力工程定额标准与市场价格波动规律,建立动态标准化体系,消除不同来源、不同时期造价数据的口径差异,为模型构建提供更优质的数据支撑。

4.3 挖掘技术与造价管理的融合优化

挖掘技术与造价管理的融合优化需打破技术与管理脱节困境,贴合电力工程造价全周期管理需求推进深度融合。将异常数据挖掘流程嵌入造价管理各环节,实现数据采集、异常识别、偏差修正的一体化推进,提升造价管理的及时性与针对性^[5]。优化挖掘结果的应用模式,将异常数据识别结果与造价管控流程联动,梳理异常数据关联的管理短板,为造价管控策略调整提供数据支撑。结合造价管理数字化趋势,推动挖掘技术与造价管理信息系统深度对接,实现异常数据的实时识别与动态预警,推动造价管理从被动整改向主动防控转变,提升造价管理整体效能。

结束语

基于决策树的电力工程造价异常数据挖掘技术,通过适配逻辑分析、技术路径构建与优化方向探索,为造价管理提供了有效手段。从决策树模型优化到数据预处理技术完善,再到挖掘技术与造价管理的深度融合,全方位提升了异常数据挖掘的精准度与实用性。该技术的应用有助于及时发现造价数据中的异常,为造价管控策略调整提供依据,推动电力工程造价管理向更科学、高效的方向发展,切实保障电力工程项目的经济效益。

参考文献

- [1]邱妍.基于决策树的电力工程造价异常数据挖掘技术[J].信息记录材料,2026,27(4):114-116.
- [2]邓嘉翕,李小云,周松林,等.基于改进随机森林的电力工程造价异常数据挖掘方法[J].信息技术,2025(5):94-98.
- [3]邵帅,赵祥,敖慧凝,等.基于梯度提升决策树算法的电力工程造价预测模型[J].沈阳工业大学学报,2025,47(3):302-308.
- [4]何鹏.变电站项目造价影响因素与科学控制途径分析[J].工程技术研究,2021(7):122.
- [5]邓佳.电力工程造价控制中控制策略分析[J].集成电路应用,2021(7):162-163.