

基于客户档案多模态关联分析的电力营销数据全生命周期溯源方法

张多志¹ 胡晓祥¹ 梁锐凯¹

ZHANG DuoZhi HU Xiaoxiang LIANG Ruikai

摘要

为提高数据溯源中对关键信息的识别能力及溯源查询效率,文章基于客户档案多模态关联分析,以电力营销数据为例,设计其全生命周期溯源方法。首先,针对客户档案多模态数据,通过统一表达与分类处理,建立联合协作分析模型,构造目标函数以降低多模态数据输出误差,提高分类精度,并实现同类别数据聚类,完成数据分类处理;其次,利用数据反演技术,通过属性反演和反演重建数学模型进行属性逆向还原,结合投影反演重构,实现电力营销数据全生命周期源头数据识别;最后,对溯源信息进行形式化描述,建立适用于用电场景的建模方法,定义溯源模型中的数据实体与追踪链,通过功能逻辑描述追溯性数据,实现异构系统数据连接共享,进而完成电力市场数据全生命周期的跟踪与溯源。实验结果表明,该设计方法应用后,可提高数据全生命周期溯源中的特征信息选择能力,提升电力营销数据全生命周期溯源效率。

关键词

客户档案;溯源方法;全生命周期;电力营销数据;多模态关联分析

doi: 10.3969/j.issn.1672-9528.2026.06.028

0 引言

电力营销作为电网企业与用户交互的核心环节,承载着海量、异构且动态增长的用电数据。这些数据贯穿于客户服务、电费计收、负荷管理、能效分析等营销业务的全生命周期,是企业实现精准决策、提升服务质量、优化资源配置的关键资产。然而,随着电力市场改革的深化和数字化转型的加速,传统数据管理方式正面临严峻挑战:数据来源多样,格式不一,且在采集、传输、存储与应用环节中可能存在数据不一致、质量参差、血缘模糊等问题,导致难以对数据异常或业务争议进行快速、准确的根源定位与责任追溯。

郭兆华等^[1]以 $12\ 500\sim 4\ 000\ \text{cm}^{-1}$ 波段光谱数据为基础,研究光谱转换、噪声增强、组合等三种数据增强策略,采用焦点 Loss 损耗函数进行建模。但近红外光谱不能适应复杂基体样品,易受到水分、粒度等干扰,从而降低了特征提取的稳定性。但深度学习的“黑箱”特征使得该方法的解释性较差,很难达到中药溯源研究的严谨科学要求。高官岳等^[2]采用哈希算法对数据进行加密,保证了数据的完整性,使用智能合约实现了对多条件的查询,降低了人工介入的错误;采用数据切片的方法,实现了在多个节点上的分布式存储,提高了系统的并行性;提出基于公钥和私

钥的分级访问权限管理机制,以保证数据的安全性。但区块链的分布式存储机制造成了大量的数据冗余,增大了存储开销,然而智能合约中存在的代码缺陷会受到恶意攻击,从而对系统安全构成威胁。

目前多模态高维数据的相关性研究中,仅将多模态数据进行融合,难以保证高效挖掘任务的完成,而且容易产生“过拟合”现象,从而降低计算的精度。因此,如何在避免过度拟合的前提下,高效地实现多源数据的关联分析,从而从全生命周期中挖掘出电力市场营销数据的核心信息,为后续的决策和预测提供支持,是目前大数据领域的核心问题。为实现对此方面工作的深化,本文将基于客户档案多模态关联分析,以电力营销数据为例,对其全生命周期溯源方法展开设计。

1 基于客户档案多模态关联分析的数据分类处理

在对电力市场进行全生命周期追溯的过程中,客户档案作为主要信息来源。为确保数据溯源的可靠性,需要对数据进行多模态关联处理。将客户画像中的多模态数据进行统一表达,并实施分类。将文字模式资料设置成一组特定尺寸的资料集。根据客户资料的分类,将其再细分为若干子分类。通过建立多模态数据的联合协作分析模型,对多模态数据进行深度挖掘^[3]。在建立模型时,不同数据对溯源的作用,并

1. 甘肃同兴智能科技发展有限公司 甘肃兰州 730050

考虑不同类型样本间的相关性。构造基于客户档案多模态关联分析的数据分类处理目标函数^[4]。用公式表示为:

$$K = \min \|S - W \times \alpha\|_2^2 + \sum_{i=1}^i \|L_i - T_i \times \beta\| \quad (1)$$

式中: K 表示基于客户档案多模态关联分析的数据分类处理目标函数, S 表示多模态数据输出结果, W 表示文本数据矩阵, α 表示文本数据权重, i 表示类别数, L 表示各类输出结果, T 表示图像数据子矩阵, β 表示图像数据权重。通过对各模态数据进行测量, 并将其与输出的误差降至最低, 从而提高分类精度, 其后的范数项用于调控相关矢量的稀疏性, 以保证模型能抽取最有代表性的特征。

为确保分类中对相同样本的联合分析, 在完成目标函数的具体处理后, 进行同类别数据的聚类, 用公式表示为:

$$J = \sum_i K \|M_i - \delta_i\|_2^2 \quad (2)$$

式中: J 表示同类别数据的聚类, M 表示惩罚项, δ 表示数据稀疏度。通过上述方式, 能够实现基于客户档案多模态关联分析的数据分类处理。

2 电力营销数据全生命周期源头数据识别

通过对聚类后的数据进行反演, 实现全生命周期源头数据的识别。数据反演主要包括属性反演、投影反演重建和导出反演^[5]。其中, 属性反演与反演重建是辨识源数据的关键, 在电力市场中, 用户画像的多模态分类数据是追溯产品来源的重要基础, 通过一系列的变换和映射操作, 将其保存在各个数据节点上。建立基于属性反演和反演重建的数学模型, 并进行属性反向还原功能设置。通过属性逆向还原, 得到电力营销数据中的源头数据^[6]。用公式表示为:

$$f_{-1}^a = C_k^a(J) \rightarrow g_{-1}^a(G) \quad (3)$$

式中: f_{-1}^a 表示经属性逆向还原得出的电力营销源头数据求解结果, C_k^a 表示经过属性转换后的客户档案多模态分类数据, g_{-1}^a 表示属性转换函数, G 表示逆求解规则。

上述公式由变换后的属性值反演得到初始属性值, 从而为判别源数据的属性层次提供依据。在此基础上, 进行投影反演的重构, 确保识别的全生命周期源头数据符合技术规范^[7]。用公式表示为:

$$f_{-1}^m = C_k^m(J) \rightarrow g_{-1}^m(G) \quad (4)$$

式中: f_{-1}^m 表示重构后的源头数据, m 表示映射系数。该方法利用映射后的数据, 对原数据进行重建, 并与属性反求功能协同, 实现对源数据的求解。在电力市场数据全寿命周期流转过程中, 基于各存储区块间的关联性, 通过反演的方法, 逐级确定整个电力市场数据的生成和变化过程, 实现电力营销数据全生命周期源头数据识别^[8]。

3 电力营销数据溯源

在电力营销数据溯源研究中, 为实现对源头数据的有效管理与跨系统追溯, 首先对其形式化建模。本文采用适用于用电场景的建模方法, 通过对数据结构与关系的清晰表达, 提升数据溯源能力。该方法借鉴了已有研究中针对电力市场数据的表达思路^[9], 将关键信息封装为一种特定结构——由双节点和一条有向边构成的“数据段”。该结构能够直观刻画电力数据之间较强的关联关系, 并支持异构系统间追溯信息的交互, 从而全面描述数据在全生命周期内的溯源路径。

在溯源模型中, 数据间的追溯关系通过功能逻辑进行形式化描述。具体而言, 将关系项置于主项与宾项之前, 作为函数参数以表达二者之间的相关性。该语法支持定义默认命名空间和可选前缀, 兼容通用数据序列化格式, 便于识别来自不同词表的标识符。这一机制不仅增强了模型对关联数据的灵活处理能力, 还将数据及其关系以机器可读的形式加以描述, 为实现异构系统间的数据连接与共享提供了基础^[10]。

可追溯模型中的节点有三种: 实体、活动、代理。在电力营销市场中, 可以通过公式定义数据实体。

$$B = E(\gamma, [s-h]) \quad (5)$$

式中: B 表示溯源中的数据实体, E 表示溯源链, γ 表示标识符, s 表示属性, h 表示值对集合, $s-h$ 表示集合中实体有关的描述性信息。在此基础上, 参照上述方式, 分析电力营销过程中数据的变化趋势。以此种方式, 构建数据的变化趋势链, 以便于对数据的追踪。用公式表示为:

$$b = A(B)(F, t_1, t_2, [H], [s-h]) \quad (6)$$

式中: b 表示数据的追踪链, A 表示活动节点强制标识, F 表示有效标识符, t_1 表示开始时间, t_2 表示结束时间, H 表示参数集合。

在此基础上, 与已有的供电销售数据源数据相融合, 以源数据为基础, 按照节点之间的相关性, 循着有向边缘逐级追溯, 可以对电力市场的整个生命周期进行跟踪, 并对其从生成到利用的全过程进行清晰地把握。通过上述方式, 实现基于客户档案多模态关联分析的电力营销数据全生命周期溯源方法。

4 对比实验

4.1 实验准备

本文选择某大型电力营销中心作为研究试点, 此单位服务地区若干重点供电企业。营销中心终端系统包括 1 套智能化的电能表数据采集模块、2 套用电信息管理系统、分散式数据处理模块、多层体系结构的服务器等。智能电表的数据采集模块为 15 min/次, 用电信息管理系统模块 1 h/次, 采样数据准确性达到工业 A 级。

终端系统具有 500 TB 的存储空间, 服务器配置备用供

电与散热系统。对各部件性能指标开展仿真计算，测得数据采集模块使用寿命为8年，存储装置为10年，网络装置为6年。将终端系统接入测试设备，搭建如图1所示的测试环境。



图1 电力营销数据全生命周期溯源测试现场

为确保溯源信息的可靠性，排除实验中相关因素的影响，按照表1，进行测试终端技术参数的配置。

表1 测试现场技术参数的配置

序号	项目	参数
1	测试平台	Ubuntu
2	编译器	GoLand
3	数据容器	Docker
4	开源区块链平台	Hyperledger Fabric
5	数据库	MySQL 7.1

为确保实验结果的可比性，引进文献[1-2]提出的溯源方法，将其作为对照，进行电力营销数据全生命周期溯源的相关测试。

4.2 数据全生命周期溯源中的特征信息选择能力检验

数据全生命周期溯源中的特征信息选择能力检验，侧重于评价追溯法对整个电力市场营销数据的选择能力。以假阳性率作为横轴，真阳性率作为纵轴，建立坐标系，画出对应的曲线。横坐标是指把不重要的特性错误地当成重要特性的可能性，而纵坐标则是指对关键性特性的正确辨识。

以区域范围为核心判断标准，大范围内的溯源算法能够从海量的市场营销大数据中准确提取出有价值的特征，从而更好地捕捉到复杂变化的全生命周期中的关键特征，为进一步的溯源分析提供更加可信的信息支持。以此为依据，对实验结果进行分析，如图2所示。

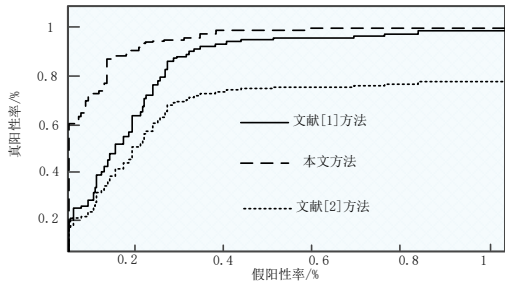


图2 数据全生命周期溯源中的特征信息选择能力检验结果

从图2所示的实验结果可以看出，应用本文方法进行电力营销数据全生命周期溯源，该方法应用后的区域面积最大，

说明该方法在溯源中的数据全生命周期溯源中的特征信息选择能力较强。

反之，文献[1-2]方法应用后的区域面积相对较小，说明对应方法在溯源中对特征信息的选择能力较弱，容易出现溯源结果偏差等问题。

4.3 全生命周期溯源效率检验

该指标的目的在于研究当数据规模发生动态变化时，研究电力市场销售数据的生命周期跟踪方法的有效性。着重考察了在整个寿命周期内，随着数据的增加，追溯过程所需要的时间变化。

随着电力市场的不断发展，其数据将随着时间的推移而不断累积，追溯的效率将直接关系到商业决策的时效性。

通过对不同规模的数据进行跟踪，分析其随数据增加的规律，明确跟踪算法对海量数据的处理能力，并判定其能否适应电力市场对可追溯性的要求，保证商业过程的流畅运行。实验结果如图3所示。

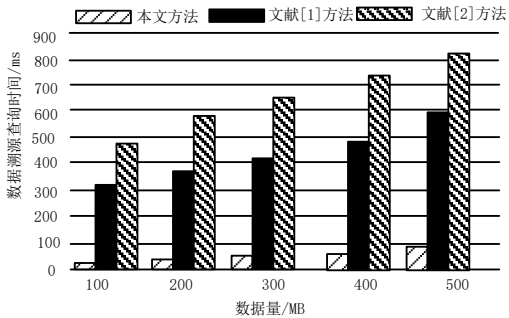


图3 电力营销数据全生命周期溯源效率检验

从图3所示的结果可以看出，随着全生命周期数据量的增加，电力营销数据溯源所需要的时间呈现出递增趋势。但在进一步分析实验结果中发现，在相同的测试条件下，应用本文方法进行电力营销数据全生命周期溯源，数据溯源查询所需要的时间一直稳定在<100ms范围内。

而文献[1-2]方法的数据溯源查询时间则相对较高，且随着数据量的增加，数据溯源查询的最长时间>800ms，已存在明显的延迟。

5 结语

本文通过基于客户档案多模态关联分析的数据分类处理、电力营销数据全生命周期源头数据识别、电力营销数据溯源，完成了此次设计，为提高数据的可信度提供科学依据。但是，受限于时间和调查对象的局限性，数据的溯源模式解析还不够深入，本体的构建还需要改进，组织本体还需要进一步提炼，工具本体需要更多的细节提高其自适应能力。在后续的工作中，将结合相关工作的具体需求，探索质量评价机制和质量溯源计划，为电力行业的进一步发展提供技术支持。

火电厂多源异构数据融合的设备安全态势 AI 感知方法

许 枫¹ 王晓峰¹ 郑锁成¹
XU Feng WANG Xiaofeng ZHENG Suocheng

摘 要

火电厂复杂电磁环境下, 设备状态监测面临多源异构数据高维冗余、特征耦合性强、状态突变点难以精准捕获, 以及强噪声干扰下设备早期劣化趋势感知响应迟滞等问题, 文章提出一种基于多尺度特征融合与动态自适应分割的设备安全态势 AI 感知方法。构建改进层次神经网络与决策树融合模型, 通过文本化转换与卷积特征提取实现异构数据降维与深度特征融合。引入视觉分析 AI 算法, 结合高斯滤波与多特征加权融合机制, 优化图像目标的纹理与边缘特征提取精度。采用累积和算法实现运行参数时间序列的动态分割与突变点检测, 并构建多时间尺度趋势融合模型, 引入安全态势长度指数量化设备安全裕度。实验结果表明, 所提方法在变压器线路监测中态势感知误差小于 20 MVA, 数据融合耗时低于 2.5 s, 能够有效预警设备异常, 提升火电厂设备安全监测的准确性与实时性。

关键词

火电厂; 多源异构数据; 设备; 安全态势; 感知

doi: 10.3969/j.issn.1672-9528.2026.06.029

0 引言

火力发电厂作为国家电力供应的核心支柱, 其安全稳定运行直接关系到社会经济发展。火电厂在实际运行中面临着设备老化等多种安全风险。为实现火电厂的安全可靠运行,

必须对海量的多源异构数据进行高效融合和智能分析。在研究中, 陈明等^[1]建立了统一威胁情报管理系统, 通过制定个性化响应策略及利用自动化工具支持, 实现快速响应与情报反馈, 从而优化模型准确性。但当前情报收集存在滞后性, 部分情报来源的可靠性不足, 影响了感知准确性。林广朋等^[2]构建了径向基函数神经网络, 通过非线性映射捕捉攻击行为

1. 国能神皖合肥发电有限责任公司 安徽合肥 230000

参考文献:

- [1] 郭兆华, 文师召, 李思凡, 等. 基于近红外光谱结合数据增强 CNN 算法的白芷产地溯源方法 [J]. 中国药学杂志, 2024, 59(21):2022-2029.
- [2] 高官岳, 孙传恒, 罗娜, 等. 基于区块链的农产品供应链溯源数据多条件查询优化方法研究 [J]. 农业机械学报, 2024, 55(3):362-374.
- [3] 王璠, 汪江, 温炜, 等. 映射关系下多模态微电网数据实时检索仿真 [J]. 计算机仿真, 2024, 41(11):173-177.
- [4] 朱铜, 艾松, 陈琨, 等. 应用数据基线模型的重燃叶片制造质量异常溯源路径搜索方法 [J]. 西安交通大学学报, 2024, 58(2):136-145.
- [5] 李艳华. 跨境数据流动中欧美间政府访问数据的理论溯源、现实矛盾与监管发展 [J]. 情报杂志, 2023, 42(7):86-92.
- [6] 陈冬英, 张昊, 张子龙, 等. 基于改进 1D-VD-CNN 与近红外光谱数据的金银花产地溯源研究 [J]. 光谱学与光谱分析, 2023, 43(5):1471-1477.
- [7] 孙传恒, 魏玉冉, 邢斌, 等. 基于智能合约和数字签名的

马铃薯种薯防窜溯源研究 [J]. 农业机械学报, 2023, 54(7): 392-403.

- [8] 孙传恒, 袁晟, 罗娜, 等. 基于区块链和边缘计算的水稻原产地溯源方法研究 [J]. 农业机械学报, 2023, 54(5): 359-368.
- [9] 田超, 郑皎凌. CTS: 基于拥堵溯源算法的信号灯多智能体强化学习组织方案 [J]. 计算机应用研究, 2023, 40(1): 178-184.
- [10] 潘华良, 丁海斌, 唐密, 等. 公路建设项目多维度档案数据溯源管理系统的作用与功能实现 [J]. 北京档案, 2023(5): 34-37.

【作者简介】

张多志 (1971—), 男, 甘肃张掖人, 本科, 工程师, 研究方向: 电力营销、市场拓展。

胡晓祥 (1986—), 男, 甘肃武威人, 本科, 中级工程师, 研究方向: 电气工程。

梁锐凯 (1992—), 男, 甘肃兰州人, 本科, 助理工程师, 研究方向: 计算机科学与技术。

(收稿日期: 2025-10-10 修回日期: 2026-05-28)