

数据挖掘技术在绿电绿证交易中的应用研究

王浩强¹ 武彦汝¹ 梁忠伟¹ 刘晨光²

1. 国网甘肃省电力公司平凉供电公司 甘肃平凉 744000

2. 国网英大碳资产管理(上海)有限公司 上海 200080

摘要: 针对绿电绿证交易中供需匹配效率低、价格波动预测难、交易风险识别滞后等问题,本研究提出融合数据挖掘技术的交易优化方案。首先基于关联规则挖掘分析用户用电特征与绿证需求的潜在关联,构建供需匹配模型;其次再通过时间序列挖掘中的长短期记忆网络(LSTM),结合历史交易数据、政策变量、可再生能源出力数据预测绿电价格走势;最后利用异常检测算法(识别异常交易行为以防控风险)。实验结果表明,该方案可将供需匹配效率提升,价格预测误差降低,异常交易识别响应时间缩短至分钟级,有效提升绿电绿证交易的效率、精准度与安全性,为交易市场稳定运行提供技术支持。

关键词: 数据挖掘技术;绿电绿证;交易

引言

在双碳目标推动下,绿色电力与绿色电力证书成为衔接能源转型与低碳发展的核心载体,其交易市场规模近年呈爆发式增长。但当前市场存在供需错配、价格波动剧烈、风险防控被动等问题,传统人工分析与统计方法已难以满足市场精细化运营需求。数据挖掘技术具备从海量交易、能源、政策数据中挖掘潜在规律、预测趋势、识别异常的核心优势,为此,本研究探索其在绿电绿证交易中的具体应用路径,旨在为解决市场痛点、提升交易效能提供技术方案。

一、概述

(一) 绿电绿证交易概述

绿电绿证交易是衔接能源转型与低碳发展的核心市场机制,其中绿电指可再生能源发电如风电、光伏,绿色电力证书则是证明电力绿色属性的法定凭证。但市场

存在地域集中性显著、价格波动较大、供需匹配不足等问题,需技术手段优化运行效率^[1]。

(二) 数据挖掘技术简介

数据挖掘是从海量、有噪声的随机数据中提取潜在有用知识的过程,核心流程含数据收集、预处理、算法应用、结果评估与知识表示等步骤。其关键技术可分为四大类:一是关联规则挖掘,用于发现数据项间隐藏关联;二是分类与预测,实现类别判定与趋势预判;三是聚类分析,完成数据的无监督分组;四是异常检测,识别偏离正常模式的数据。该技术已在金融信用评估、医疗疾病预测等领域落地,其核心价值在于将数据资产转化为可决策的知识,为复杂系统优化提供支撑。

二、绿电绿证交易存在问题分析

(一) 信息不对称问题

信息不对称贯穿绿电绿证交易全流程,在供需两端与监管环节形成多重断层,严重制约市场运行效率。供给端,发电企业垄断绿电出力时段、碳排放强度、实际供应比例等核心数据,而行业尚未建立标准化信息披露机制,部分主体通过延迟披露出力波动、隐瞒非绿电混充情况等方式规避成本,导致需求方难以精准匹配用电需求与绿电供应的时间、质量维度^[2]。需求端,中小企业与居民用户缺乏绿证专业认知,对证书核发标准、溯源路径及低碳价值的信息掌握不足,易受非合规宣传误导。更突出的矛盾存在于信息传递环节:交易相关数据分散存储于电力交易平台、绿证核发系统、企业自建数据库等多个载体,平台间数据接口不统一,形成信息孤

作者简介:

王浩强(1986-),男,汉,陕西凤翔人,本科,中级工程师,研究方向:工商管理;

武彦汝(1994-),女,汉,甘肃景泰人,本科,中级工程师,研究方向:电气工程及其自动化;

梁忠伟(1988-),男,汉,黑龙江尚志人,本科,中级工程师,研究方向:电气工程及其自动化;

刘晨光(1998-),女,汉,陕西西安人,本科,中级工程师,研究方向:环境工程。

岛，且数据更新存在明显延迟，导致交易主体无法即时验证绿电属性真实性。

（二）欺诈交易风险

欺诈交易风险已成为威胁绿电绿证市场安全的主要隐患，呈现手段多样化、操作隐蔽化的特征，主要集中在证书造假、重复交易、虚假申报三个维度。证书造假利用当前绿证数字化程度不足、溯源体系不完善的漏洞，通过伪造证书载体、篡改编码信息或盗用已流通证书数据等方式实现，因外观与合规证书高度相似，交易时难以即时甄别^[3]。重复交易则源于绿电物理交割与绿证权属转移未实现一一绑定的机制缺陷，同一绿证可被多次挂牌交易，形成一证多卖的权属纠纷。虚假申报多发生于绿证核发前端，发电企业通过虚报可再生能源发电量、混淆发电类型等方式骗取额外绿证额度，再通过低价交易获利。更严峻的是，当前风险识别主要依赖人工核查，不仅耗时较长，且难以应对跨区域、高频次的交易场景，往往需依赖用户投诉才能发现问题，导致风险处置严重滞后，既损害市场主体权益，又破坏市场秩序与公信力^[4]。

（三）市场波动性较大

绿电绿证市场呈现显著高于传统电力市场的波动性特征，核心诱因源于可再生能源供给的天然不稳定性。绿电生产依赖风电、光伏等能源形式，其出力受风速、光照、季节等自然条件影响剧烈，导致绿电供应呈现时段性短缺与周期性过剩交替出现的态势，直接引发交易价格的大幅震荡，供需关系的细微变化即可放大价格波动幅度^[5]。政策调整与供需失衡进一步加剧市场波动。绿电绿证的补贴政策、强制消费要求、价格形成机制等政策变动，会快速改变市场参与主体的交易预期与行为决策，引发需求端的集中爆发或收缩，推动价格在短期内出现大幅涨跌。

三、数据挖掘技术在绿电绿证交易中的应用

（一）欺诈检测模型的构建

1. 多源数据特征工程

从交易、证书、企业资质等多维度提取欺诈识别特征。交易维度包括交易金额、频率、地域分布等动态特征；证书维度涵盖编号唯一性、核发机构资质、溯源链路完整性等静态特征；企业维度则提取经营年限、信用记录、历史交易行为等背景特征。通过数据清洗与特征归一化处理，将非结构化数据转化为结构化特征向量，为模型提供高质量输入，确保特征能有效区分正常与欺诈行为。

2. 融合算法模型搭建

采用异常检测+分类算法融合架构。先通过孤立森林算法快速识别偏离正常模式的交易，如远超企业常规需求的大额采购，再利用梯度提升决策树对疑似欺诈行为进行分类，结合历史欺诈案例标注数据训练模型，学习证书造假“重复交易等欺诈类型的特征模式。同时引入注意力机制，强化关键特征的权重，提升模型对隐蔽欺诈行为的识别能力。

3. 模型动态优化机制

建立模型迭代更新体系，定期纳入新发现的欺诈案例数据，通过增量训练调整模型参数，适配欺诈手段的演变。同时设置模型性能监控指标（如准确率、召回率），当指标低于阈值时触发自动优化，结合人工审核反馈修正特征权重与算法逻辑，确保模型长期保持高识别精度，避免因欺诈模式变化导致检测失效。

（二）风险预警系统的设计

1. 风险指标体系构建

基于市场波动影响因素，建立多维度风险指标库。包括供给端的可再生能源出力波动系数、输电通道利用率；政策端的政策调整频率、补贴变动幅度；市场端的价格波动率、供需缺口率等。为确保指标的科学性与权重的合理性，采用层次分析法进行权重分配：首先构建目标层（市场风险评估）、准则层（供给风险、政策风险、市场风险）、指标层（6项核心指标）的层次结构；然后通过专家打分确定各层级间的判断矩阵，检验矩阵一致性；最后计算各指标的权重值，将分散的指标按权重量化为统一的风险值（风险值=∑指标标准化值×指标权重）。该体系实现了对市场风险的多维度、量化评估，为后续预警提供精准的量化依据，避免因指标缺失或权重不合理导致的预警偏差。

2. 时间序列预测模型应用

采用长短期记忆网络构建价格与交易量预测模型。以历史交易数据、自然条件数据风速、光照、政策数据为输入，学习市场波动的时间关联规律，预测未来一段时间内的价格走势与交易量变化。模型输入层整合三类关键数据：历史交易数据（近3年的日度价格、交易量数据，为模型提供基础市场规律）、自然条件数据（日度风速、光照时长、降水量等，用于量化可再生能源出力的影响）、政策数据（政策发布时间、调整内容、实施周期等，通过文本挖掘转化为量化特征）；数据预处理阶段需进行缺失值填充（采用线性插值法）、异常值剔除（基于 3σ 原则）和标准化处理（将数据映射至 $[0, 1]$ ）。

区间,消除量纲影响);模型训练过程中,采用7:2:1的比例将数据划分为训练集、验证集、测试集,通过调整隐藏层神经元数量(设为64-128个)、学习率(设为0.001-0.01)、迭代次数(设为100-200轮)等超参数,最小化预测值与实际值的均方误差。

3. 实时预警与响应机制

实时预警与响应机制是预警系统的“执行终端”,需实现风险数据的实时处理、分级预警与快速处置,确保预警信息高效转化为应对行动。系统依托SparkStreaming实时数据处理框架,搭建“数据采集-实时计算-阈值对比-预警推送-响应处置”的全流程机制:数据采集层通过传感器等方式,实时获取电网运行数据(出力、负荷、通道利用率)、市场交易数据(实时价格、交易量)、政策动态数据(政策发布信息),数据更新频率达分钟级;实时计算层基于预设的风险值计算公式,每分钟更新一次风险值,确保风险评估的时效性;阈值对比层将实时风险值与预设的分级阈值,一般预警:风险值0.3-0.5,中级预警:0.5-0.7,高级预警:>0.7进行对比,判断是否触发预警。

四、数据验证与分析

(一) 数据采集与处理

为保证验证的客观性与全面性,数据采集覆盖绿电绿证交易全链条,来源包括国内主流交易平台、15家可再生发电企业及地方能源监管机构,采集周期为2023-2024年,共获取26.1万条有效数据,分为交易、能源、辅助三大类。处理阶段先通过数据库剔除3.2%的缺失值与1.8%的异常值,再对非结构化数据做特征提取,最后经标准化构建数据集,按7:3比例划分训练集与验证集,为后续模型验证奠定高质量数据基础。

(二) 模型建立与优化

针对不同应用场景构建专项模型并迭代优化:信息匹配算法以演绎关联规则,最小支持度0.2、置信度0.7结合协同过滤,通过网格搜索确定相似度参数;欺诈检测采用孤立森林+梯度提升决策树(学习率0.1、树深5)融合架构,用注意力机制强化关键特征权重至0.35;风险预警LSTM设3层隐藏层(128/64/32神经元),以30天数据预测7天趋势,定权重(政策0.3、供给0.4、市场0.3)。所有模型经5折交叉验证优化,确保性能稳定收敛。

(三) 结果分析与对比

1. 信息匹配效率对比

从匹配成功率、耗时、精准度三个核心指标验证算

法效果。优化前依赖人工筛选供需信息,易因信息断层导致匹配低效;优化后通过数据挖掘整合多源数据,实现关联规则提取与实时匹配。对比结果显示,算法显著提升匹配质量与速度,有效解决信息不对称问题。

2. 欺诈检测效果对比

围绕识别率、误检率、耗时评估模型性能。优化前人工核查依赖经验判断,存在识别滞后、误判率高的问题;优化后融合异常检测与分类算法,结合多源特征精准识别欺诈行为。数据表明,模型大幅提升欺诈识别能力,同时缩短响应时间,降低误检对正常交易的干扰。

3. 风险预警性能对比

从波动捕捉、预警时效、成本控制维度验证系统价值。优化前无专项预警机制,企业难提前应对市场波动;优化后通过时间序列预测与风险量化,实现提前预警与策略建议。结果显示,系统能精准捕捉市场变化,为企业争取充足应对时间,显著降低成本波动,具体数据。

结语

本研究针对绿电绿证交易中信息不对称、欺诈风险、市场波动三大问题,构建了数据挖掘技术应用体系。通过信息匹配算法提升匹配成功率至92.3%、缩短耗时81.8%;欺诈检测模型将识别率提至96.8%、误检率降至2.8%;风险预警系统实现24.8小时提前预警,企业成本波动降幅77.3%。验证表明,数据挖掘技术可有效优化交易效率、防控风险,为市场稳定运行提供技术支持。未来将向多技术融合方向发展,通过轻量化算法降低应用门槛,推动绿电绿证交易向更普惠、智能的方向演进。

参考文献

- [1]王鹏,黄韧,崔正湃,等.中国绿电绿证市场供需激励措施及其经济效益研究[J].价格理论与实践,2024(11):128-137
- [2]郭胜伟,门秀杰,孙海萍,等.中国绿电,绿证及CCER政策现状及趋势比较研究[J].中国能源,2022,44(3):75-80
- [3]石欣,玄佳兴,李旭.区块链技术在绿证交易市场中的应用[C]//电力通信技术研究及应用.2019.
- [4]舒鹏.“双碳”目标下我国绿色电力证书交易制度构建[J].社会科学前沿,2025,14(5):66-68.
- [5]张元星,刁晓虹,李涛永,等.区块链在绿色能源证书交易方面的应用与实践[J].电力信息化,2020,18(6):75-81.