DOI:10.20033/j.1003-7241.(2026)07-0180-04

基于随机森林算法的电费回收风险自动预警系统

黄蓓雯, 陈冰旭

(国网上海市电力公司客户服务中心, 上海 200030)

摘　要:针对现有电费回收风险自动预警系统在风险等级划分上缺乏有效机制、难以对风电等波动性较强的能源类型进行精确的风险等级划分,导致整体预警精度不足的问题,提出了一种基于随机森林算法的电费回收风险自动预警系统。通过客户欠费风险预测子系统、用户管理子系统、客户信用评级子系统、风险管理预警子系统和行业分析预测子系统构成预警系统整体框架。构建电费回收风险指标体系,并利用熵权法对电费回收风险指标的权重进行计算,根据权重计算结果删除弱影响指标。采用随机森林算法对电费回收风险等级进行划分,实现电费回收风险的自动预警。实验结果表明,所提方法具有较高的预警覆盖率、命中率及预测精度,能够显著提升电费回收风险防控能力。

关键词:随机森林算法;电费回收风险;预警系统;风险等级;风险指标;熵权法

中图分类号: TM73;TP273　　**文献标志码**: A　　**文章编号**: 1003-7241(2026)07-0180-04

Automatic early warning system of electricity fee recovery risk based on random forest algorithm

Huang Beiwen, Chen Bingxu

(State Grid Shanghai Municipal Electric Power Company Customer Service Center, Shanghai 200030, China)

Abstract: Aiming at the issue that existing automatic early warning systems for electricity fee recovery lack effective mechanisms for risk level classification and struggle to accurately classify risk levels for highly volatile energy types such as wind power, resulting in insufficient overall warning accuracy, this paper proposes an automatic early warning system for electricity fee recovery based on the random forest algorithm. The whole framework of the early warning system is composed of the customer default risk prediction subsystem, user management subsystem, customer credit rating subsystem, risk management early warning subsystem and industry analysis and prediction subsystem. The paper constructs the risk index system of electricity charge recovery, calculates the weight of the risk index of electricity charge recovery by entropy weight method, and removes the weak influence index according to the weight calculation results. The random forest algorithm is used to divide the risk level of electricity recovery to realize the automatic early warning of electricity recovery risk. Experimental results show that the proposed method exhibits high early warning coverage, hit rate, and predictionaccuracy, significantly enhancing the risk prevention and control capability of electricity fee recovery.

Keywords: random forest algorithm; risk of tariff recovery; early warning system; risk level; risk indicators; entropy weight method

电力行业在市场经济飞速发展的背景下开始加快体制改革的步伐,电力企业属于服务型企业,是社会中各个行业中企业进行一系列业务的基础[1]。正是因此独特的重要作用,在我国经济新常态下电力企业的社会责任不断加重,而电费回收风险对于电力企业来说是影响其业务开展的主要因素。电费回收的主要风险来源于服务的企业大用户[2],大用户在行情不断变化、激烈的市场竞争下存在随时倒闭的风险,在这种情况下,电力系统需要构建电力回收风险预警系统,降低电费回收风险。大多数国家随着市场经济的发展开始研究风险预警系统,并取得了一些成就,将其应用在电力行业中,为供电企业降低风险发生概率提供了相关依据。

文献[3]提出基于长短期记忆网络算法的电费回收风险预警方法,文中首先综合电力客户的电量、电费数据,以及电力客户在多个部门的风险信息,建立电力大客户的电费回收风险指标体系。其次,基于熵值法得到的风险指标权重系数,过滤弱影响指标,采用相关性分析剔除重叠作用指标,得到客户电费回收风险预警指标。最后,基于深度学习中的长短期记忆(long short-term memory,LSTM)网络算法进行了客户电费回收风险预警。但该方法存在运行耗时长的问题。文献[4]提出基于电力大数据的电力企业电费回收风险预警模型,该方法基于电力大数据确定电费回收风险预警等级约束,根据电费回收风险的内部外部因素,提出相应的预警约束条件。优化电力企业电费

收稿日期:2025-07-12;**录用日期**:2025-09-11

作者简介:黄蓓雯(1982—),女,经济师,研究方向为电力营销。

引用本文:黄蓓雯, 陈冰旭. 基于随机森林算法的电费回收风险自动预警系统[J]. 自动化技术与应用, 2026,45(7):180-183.(Huang Beiwen, Chen Bingxu. Automatic early warning system of electricity fee recovery risk based on random forest algorithm[J]. Techniques of Automation and Applications, 2026,45(7):180-183.)

回收风险预警指标权重,更加精准地衡量各个回收风险因素的重要性,确保最终的预警有效性,该方法评价过程中的弱影响指标较多,导致方法的预测精度低。

为了解决上述方法中存在的问题,提出并设计基于随机森林算法的电费回收风险自动预警系统,采用随机森林算法对电费回收风险等级进行有效划分,从而提高预警精度。

1　系统架构设计

此次所设计的预警系统由 5 个子系统构成:客户欠费风险预测子系统、用户管理子系统、客户信用评级子系统、风险管理预警子系统和行业分析预测子系统。电力回收风险自动预警系统的整体框架如图 1 所示。

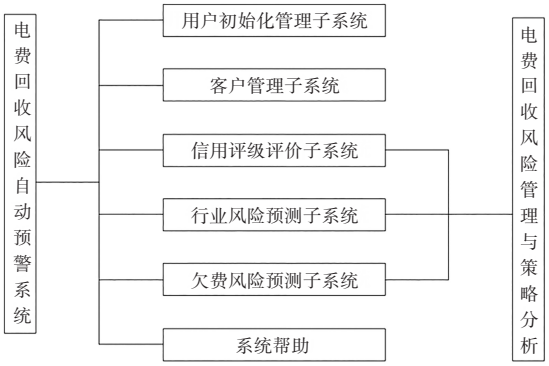


图 1　系统整体框架

Fig. 1　Overall framework of the system

根据行业分类对电费回收风险自动预警系统中的客户管理子系统进行标准化的统一管理;行业分析预警子系统在预警系统中主要负责分类电力用户,通过趋势棒图和线图反映欠费标准化数据和行业电量走势等,并预测行业风险;在客户欠费风险预测子系统中设立电费回收风险预测指标体系,预测用户欠费的风险;风险管理预警决策子系统在电费回收风险预警中的主要功能是结合欠费风险预测模型、信用等级评价模型和行业风险预测模型构建欠费风险预警决策模型,为供电企业提供服务依据,规避和防范用户欠费风险。

2　电力回收风险自动预警

2.1　指标体系构建

所设计预警系统在违约用电、电费和电量等数据的基础上,结合法院信息、税务信息和工商信息,构建电费回收风险指标体系[5-7],如图 2 所示。

2.2　指标权重计算

基于随机森林算法的电费回收风险自动预警系统采用熵值法对电费回收风险指标进行计算,具体过程如下。

设 x_{ij} 代表第 i 个电力用户的第 j 项电费风险回收风险指标对应的观测数据; P_{ij} 代表第 i 个电力用户在第 j 项电费风险回收风险指标下的特征比重,其计算公式为

$$P_{ij}=\frac{x_{ij}}{\sum_{i=1}^{n}x_{ij}} \tag{1}$$

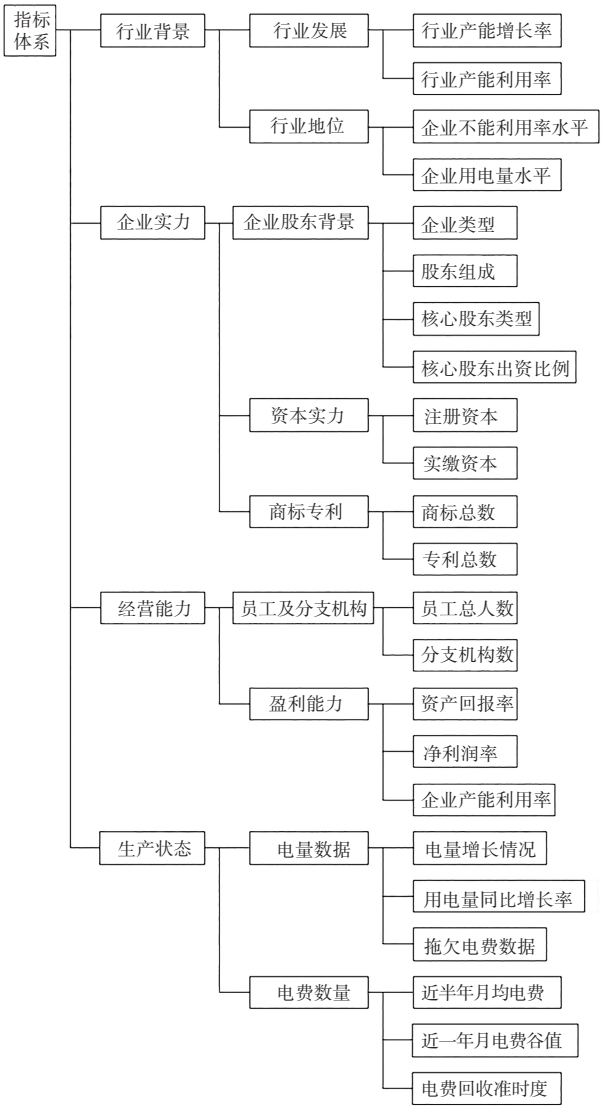


图 2　电费回收风险指标体系

Fig. 2　Electricity fee recovery risk indicator system

如果观测数据 x_{ij} 的值小于零,需要修正该观测数据。设 e_j 代表第 j 项电费风险回收风险指标对应的熵值,其表达式为

$$e_j=-k\sum_{i=1}^{n}P_{ij}\ln P_{ij} \tag{2}$$

式中, k 为常数,受电力用户总数的影响[8]。

差异系数 g_j 决定电费回收风险指标的信息效用价值,其表达式为

$$g_j=1-e_j \tag{3}$$

计算指标对应的价值系数是熵权法计算权重的本质,电费回收风险指标的价值系数与贡献率之间呈线性关系,第 j 项电费回收风险指标的权重 h_j 为

$$h_j=\frac{g_j}{\sum_{j=1}^{n}g_j} \tag{4}$$

2.3 风险预警

完成指标权重计算后,采用随机森林算法对电费回收风险等级进行划分,根据风险等级实现风险自动预警。

设定训练集合为 D,由 n 个样本构成,通过有放回随机抽样方法在训练集合 D 中获得样本训练集 D_i。用 F 描述通过节点划分获得的属性总量,通过属性总量 F 以构建决策树 T_i,并根据信息参数从分类属性中筛选出最佳属性[9]。

设 P_i 代表第 i 个样本在当前样本集合 D 中所占的比例,Entropy(D) 代表集合 D 对应的信息熵,C 为项目总数,其表达式如下:

$$\text{Entropy}(D) = -\sum_{i=1}^{c} h_j \log_2 P_i \tag{5}$$

样本集合 T 根据特征 A 进行分割处理,获得 k 个部分,此时信息熵 Entropy(D_A) 可通过下式计算得到

$$\text{Entropy}(D_A) = -\sum_{i=1}^{c} h_j \log_2 P_i \tag{6}$$

在信息熵 Entropy(D_A) 的基础上计算信息增益 Gain(D,A)

$$\text{Gain}(D,A) = \text{Entropy}(D) - \text{Entropy}(D_A) \tag{7}$$

用 SplitEntropy(D,A) 描述信息值,其计算公式为

$$\text{SplitEntropy}(D,A) = -\sum_{i=1}^{c} \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|} \tag{8}$$

设 GainRation(D,A) 代表信息增益率,其计算公式为

$$\text{GainRation}(D,A) = \frac{\text{Gain}(D,A)}{\text{SplitEntropy}(D,A)} \tag{9}$$

根据信息增益率计算结果,采用随机森林算法对电费回收风险进行等级划分,具体过程如下。

1) 设定测试样本 X_{test} 和样本集 $S = \{(X_i, Y_i), i = 1,2,\cdots,n\}$。

2) 采用 bagging 方法抽样处理训练集,获得训练子集 $S_j (j = 1,2,\cdots,N)$,训练子集 S_j 即为第 j 棵决策树构成的样本集。

3) 在样本属性中随机选取 m 个属性,并对其 Gini 系数进行计算。

4) 通过最小 Gini 系数的指数对节点进行分裂处理,获得 N 棵决策树[10]。

5) 在完成训练的 N 棵决策树 T_j 中导入权重较高的电费回收风险指标集 X_{test},得到随机森林的分类结果 $F(X_{\text{test}})$,即电费回收风险预警等级划分结果为

$$F(X_{\text{test}}) = \text{GainRation}(D,A)\{T_j(X_{\text{test}})\}_{j=1}^{N} \tag{10}$$

3 实验与分析

为了验证基于随机森林算法的电费回收风险自动预警系统的整体有效性,需要对系统性能进行测试,以某电力公司一年的用户为实验样本数据,对所设计系统、文献[3]系统和文献[4]进行测试。

三种系统对电费回收风险预警的覆盖率如图3所示。

对图3进行分析可知,所设计系统在不同月份对电力回收风险预警的覆盖率均在80%以上,文献[3]和文献[4]系统在不同月份对电力回收风险进行预警时,覆盖率波动较大,表明这两种方法无法稳定地对电力回收风险进行预警。对比上述方法的测试结果可知,所设计系统对电费回收风险进行预警的覆盖率较高,因为所设计系统采用熵权法对电力回收风险指标的权重进行计算,选取贡献率高的指标实现电费回收风险的预警,进而提高了预警覆盖率。

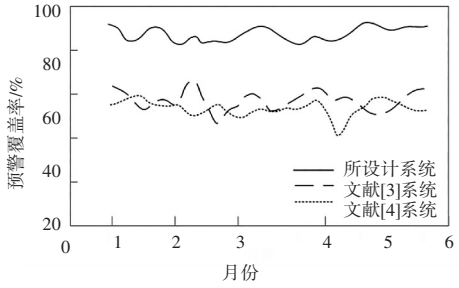


图3 预警覆盖率

Fig. 3 Early warning coverage rate

将预警命中率作为测试指标,对所设计系统、文献[3]系统和文献[4]系统的整体有效性进行测试,测试结果如图4所示。

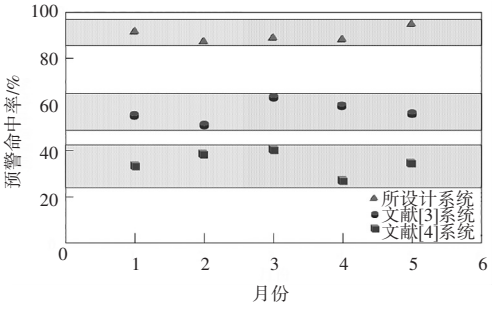


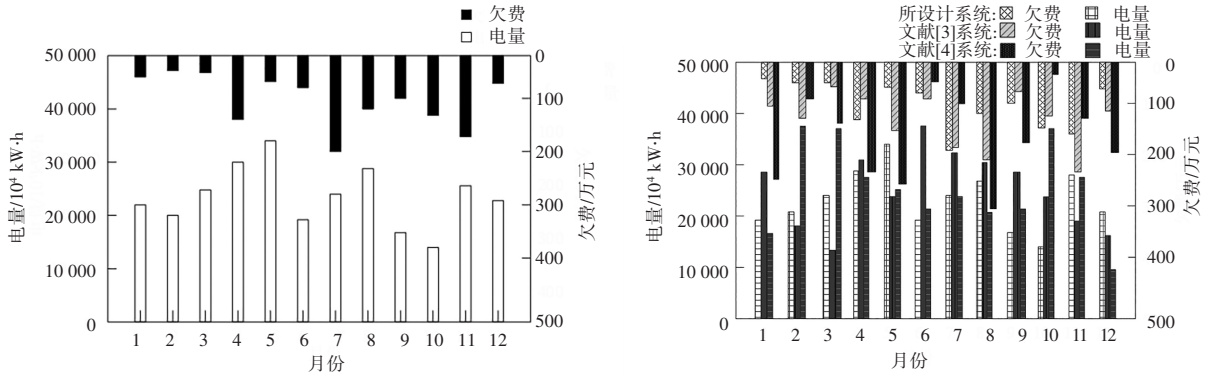
图4 预警命中率

Fig. 4 Early warning hit rate

分析图4可知,相对于文献[3]系统和文献[4]系统,采用所设计系统对电费回收风险进行预警时,预警命中率远高于文献[3]和文献[4]预警系统命中率。因为所设计系统通过计算电力回收风险指标的权重,删除了弱影响的电力回收风险指标,提高了预警命中率。

采用所设计系统、文献[3]系统和文献[4]系统对电力公司的总欠费和总用电量进行预测分析。

根据图5可知,所设计系统预测该电力公司的总欠费和总用电量与实际结果基本相符,存在的微小误差在可接受范围内,文献[3]系统和文献[4]系统预测该电力公司的总欠费和总用电量时,得到的预测结果与实际结果之间的误差较大,对电费回收风险预警的精度产生了较大的影响。对比可知所设计的预测精度高,是因为所设计系统删除了弱影响指标,利用贡献率较高的电费回收风险指标对电费回收风险等级进行划分,提高了预测结果的精度。



（a）实际总欠费和总用电量　　（b）不同系统的预测结果

图 5　预测结果

Fig. 5　Prediction results

4　结论

供电企业扩大再生产、实现可持续发展的基础是电力回收工作，电费回收通常具有滞后性，为了解决滞后性问题，供电企业提出了电费归仓策略，包括预付费和预购电等，上述策略在一定程度上可以防范电费回收风险，但针对电力大用户来说，用电成本是其生产经营成本，资金流动情况对其产生的影响较大，无法完全执行上述策略。电费回收工作是电网企业经营和管理过程中的工作重点，在此背景下亟需一套电费回收风险预警系统，提出基于随机森林算法的电费回收风险自动预警系统设计方法，构建了风险预警指标体系，采用随机森林算法实现风险预警的自动预测，解决了目前方法中存在的问题，为供电企业的经营管理提供了保障。

参考文献

[1]田园，王建波. 海外电力投资项目投建营一体化财务风险管控研究[J]. 水力发电，2024，50(2)：15-19.
[2]郑楷洪，杨劲锋，肖勇，等. 基于多源数据融合的电力客户电费回收风险画像构建[J]. 南方电网技术，2022，16(4)：135-142.
[3]谢林枫，钱立军，季聪，等. 基于长短期记忆网络算法的电费回收风险预警[J]. 电力工程技术，2018，37(5)：98-103.
[4]谭沛然，任建婧，杨天玥，等. 基于电力大数据的电力企业电费回收险预警模型[J]. 文渊(高中版)，2023(8)：203-205.
[5]叶飞，陈玉敏，杨超，等. 考虑信用动态演化的电力用户电费回收风险预警方法[J]. 电力系统自动化，2023，47(11)：112-120.
[6]黄荷，鲍凌翔. 基于隐马尔科夫模型的电费抄核误差校正系统[J]. 自动化技术与应用，2024，43(9)：155-158，188.
[7]张向东，胡文，薛斌，等. 基于云计算的电力营销统计报表核算系统[J]. 自动化技术与应用，2024，43(1)：112-116.
[8]李国栋，王峥，刘科研，等. 基于深度信念网络的配电网台区用户欠费风险预测[J]. 电力科学与技术学报，2023，38(5)：112-119.
[9]孙伟，刘阳. 考虑行业用电模式的企业电费逾期风险评估[J]. 电力需求侧管理，2023，25(4)：52-57.
[10]李汉巨. 基于孤立森林与聚类分析的恶意欠费用电客户识别研究[J]. 信息技术，2022(5)：55-60.

(上接第 157 页)

参考文献

[1]李腾飞，冯明磊，林国涛，等. 水敏性软岩挖方高边坡变形机制及其防治[J]. 公路，2022，67(1)：70-73.
[2]Uhlemann S，Fiolleau S，Wielandt S D B. Predictive monitoring of urban slope instabilities using geophysics and wireless sensor networks[J]. Leading Edge，2023，42(9)：634-643.
[3]费新峰，田野，赵超英，等. 基于多时相 InSAR 技术的黄河上游龙羊峡库区不稳定边坡识别与形变监测[J]. 地球科学与环境学报，2023，45(3)：578-589.
[4]虎勇，吕辉岩，李绍荣，等. MEMS 姿态传感器在边坡表面位移监测的应用研究[J]. 电子测量与仪器学报，2023，37(7)：53-61.
[5]李诚志，颜丙欣. 露天矿山边坡爆破技术优化研究[J]. 中国金属通报，2025(13)：50-52.
[6]李亚冰，蔡德所，黄江楼，等. 基于弱光栅的分段测斜算法研究[J]. 中国测试，2025，51(8)：54-61.
[7]李聪聪，王佟，赵欣，等. 边坡监测与治理技术在高寒矿区露天煤矿生态修复中的应用研究[J]. 中国矿业，2024，33(4)：122-131.
[8]杜建成，唐昌意，李松，等. 高陡土岩混合边坡稳定及变形分析与工程应用[J]. 公路，2024，69(3)：63-70.
[9]兰正悦，史雄风，周文龙. 基于递归变模式提取的管道滑坡变形信号分解方法研究[J]. 机械与电子，2025，43(4)：12-18.
[10]江跃，程琳，马春辉，等. 基于多点位移计的黄金峡水利枢纽施工期边坡实测变形特征分析[J]. 水资源与水工程学报，2023，34(3)：201-207，224.
[11]梁宏彦. 复杂条件下铁路超高边坡石方深孔台阶爆破施工技术[J]. 工程建设与设计，2024(2)：182-184.
[12]张龙，白旭光，田腾飞，等. 基于物联网的公路工程边坡智能动态监测系统研究[J]. 公路，2022，67(9)：122-127.
[13]苏金亮，黎盟，艾露，等. 基于北斗 GNSS 的边坡自动化实时监测数据处理及变形预测方法[J]. 水电能源科学，2022，40(5)：146-150.
[14]程林. 高陡边坡在线监测系统在南泥湖露天钼矿中的应用[J]. 黄金科学技术，2024，32(2)：366-376.
[15]费汉强，马增，潘懿，等. 基于实时动态全球定位测量的边坡监测预警研究[J]. 矿业研究与开发，2023，43(7)：54-59.
[16]段谟东，尹林，张桐源，等. 基于光纤光栅监测技术的边坡内部位移变化规律及分布特征[J]. 自动化技术与应用，2025，44(9)：12-17，110.
[17]刘晖，朱德康，许凌杰，等. 基于改进灰狼算法优化支持向量机的边坡位移预测[J]. 自动化技术与应用，2024，43(9)：30-33，171.