工程与管理科学
Engineering and Management Science
第 7 卷◆第 12 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

基于大数据分析的电力公司反窃电预警模型构建

杨朋凯　黄涛
国网河北省电力有限公司石家庄市藁城区供电分公司　052160

DOI:10.12238/ems.v7i12.16445

[摘　要] 窃电行为造成电力企业巨额经济损失，扰乱正常供用电秩序，更严重威胁电网安全运行与公共安全，传统依赖人工稽查与简单阈值判断的反窃电手段，存在效率低、覆盖面窄、滞后性强等弊端，随着智能电表（AMI）的普及和电力大数据平台的建设，海量、多维度的用户用电数据为精准识别窃电行为提供了新机遇，本文提出构建基于大数据分析的电力公司反窃电预警模型，研究整合用户历史用电量、日电量曲线、电压电流异常、表计事件记录、用户档案信息等多源数据，运用数据预处理、特征工程、机器学习（如随机森林、梯度提升树、孤立森林）与深度学习（如 LSTM）算法，构建多层级、高精度的窃电风险预警模型，实际案例验证，该模型能有效识别异常用电模式，显著提升窃电行为的发现率与稽查效率，降低企业损失，为电力公司实现智能化、精准化反窃电管理提供有力支撑。

[关键词] 反窃电；大数据分析；用电异常检测；机器学习；智能电表；预警模型

引言：

电能作为一种特殊的商品，其计量与收费的准确性直接关系到电力公司的经营效益与社会公平，窃电现象在全球范围内普遍存在，形式多样，从私拉乱接、绕越计量装置、篡改电表到利用高科技手段干扰计量等，手段日益隐蔽，据估计，部分国家和地区的技术线损与管理线损（含窃电）高达 10%-20%，给电力企业带来巨大经济损失，并可能导致线路过载、火灾等安全事故，影响电网稳定，传统的反窃电工作主要依赖用电检查人员的经验判断、周期性现场巡查以及对线损率突增等宏观指标的监控，这种方式主观性强、覆盖面有限、响应滞后，难以应对大规模、隐蔽性的窃电行为；近年来，随着高级量测体系（AMI）的广泛部署，电力公司能够实时、高频次地采集海量用户的用电数据（如每 15 分钟或每小时的电量、电压、电流、功率因数等），并结合用户档案、地理信息、天气数据等形成电力大数据，这为利用数据驱动的方法，构建智能化、自动化的反窃电预警系统创造了条件，本文旨在探讨如何利用大数据分析技术，构建高效、精准的反窃电预警模型，实现从“被动响应”到“主动预警”的转变。

一、数据来源与特征工程

构建有效的预警模型，高质量的数据与科学的特征工程是基础。

1.数据来源

构建反窃电预警模型的首要任务是整合多源数据，用电量数据作为核心数据源，来源于智能电表，能够提供用户分钟级、小时级乃至日级的用电量序列，这些高频率的数据记录了用户的用电行为模式，是识别异常用电的重要依据，此外电能质量数据如电压、电流、功率和功率因数等实时或准实时信息，对于检测潜在的窃电行为至关重要，比如电压或电流的异常波动可能表明存在绕越计量装置或篡改电表的行为，电表自诊断产生的表计事件记录也是直接的窃电线索，包括开盖记录、失压、失流、时钟异常以及参数修改等事件，这些记录为反窃电提供了重要参考，用户档案信息则涵盖了用户类型（居民、商业、工业）、合同容量、电价类别、用电地址以及历史违约记录等内容，有助于建立用户画像，进一步细化分析，外部数据如天气数据（温度、湿度、光照）、节假日信息以及区域经济发展水平等，则用于校正用电行为，确保模型在不同环境条件下的准确性。

2.数据预处理

为了确保后续模型训练的有效性，必须对原始数据进行充分的预处理，首先数据清洗是关键步骤之一，它包括处理缺失值、异常值和重复数据，对于缺失值，可以插补法（如均值插补、线性插补）或删除法来处理；异常值则可统计方法（如箱线图法）或基于模型的方法（如孤立森林）进行识别和修正；其次数据对齐与融合是将不同来源、不同时间粒度的数据进行时间对齐和关联，形成统一的用户数据视图，比如将每 15 分钟采集的用电量数据与每日更新的用户档案信息进行匹配，确保每个时间点的数据都包含完整的特征信息，这一过程需要考虑时间维度上的对齐，还需处理不同数据源之间的格式差异和单位转换问题，细致的数据预处理，可以有效提升数据的质量和一致性，为后续的特征工程和模型训练奠定坚实基础。

3.特征工程

特征工程是决定反窃电预警模型性能的关键环节,首先，基础统计特征是最基本的特征集，包括日/月/年用电量、平均负荷、负荷率、峰谷差、用电量标准差等，这些特征能够反映用户的总体用电情况及其稳定性；其次时序模式特征提取日电量曲线的形状特征来捕捉用户的用电习惯，如使用傅里叶变换、小波变换或动态时间规整（DTW）距离等方法来量化曲线的相似性和周期性，用电模式的周期性（如日周期、周周期）和趋势性也是重要的特征，可以帮助识别异常变化；第三，异常事件特征统计单位时间内开盖次数、失压/失流事件发生频率及持续时间来发现潜在的窃电行为；第四，对比特征计算与同类用户（同类型、同区域）的用电量、负荷曲线的差异度（如欧氏距离、余弦相似度），以及与自身历史同期用电量的偏离度，来发现异常行为；最后，衍生特征如日电量变化率、负荷波动率、电压/电流不平衡度等，进一步丰富了模型的输入特征，提高了模型的预测能力，全面的特征工程，可以显著提升模型的准确性和鲁棒性。

二、反窃电预警模型构建

本文构建多模型融合的预警框架，以适应不同窃电行为的特征。

（一）基于监督学习的分类模型

基于监督学习的分类模型适用于有经过稽查确认的窃电用户样本（正样本）和大量正常用户样本（负样本）的情况，这类模型学习已知样本的特征差异模式，来识别新的潜在窃电行为，在算法选择方面，随机森林（Random Forest）、梯度提升树（如 XGBoost，LightGBM）等集成学习算法因其能处理高维特征、对噪声鲁棒、可解释性相对较好，成为首选，此外支持向量机（SVM）和神经网络也可用于特定场景，具体流程上，首先使用标注数据训练模型，这些数据包括用户的用电量、电能质量数据、表计事件记录等多源信息，训练模型能够学习到窃电用户与正常用户之间的特征差异模式，比如窃电用户可能表现出异常的用电量波动或频繁的表计事件，训练完成后，模型输出每个用户为“窃电”或“正常”的概率，并根据设定的阈值生成预警名单，这种方法能提高反窃电工作的效率，减少误报率，确保稽查资源的有效利用。

（二）基于无监督/半监督学习的异常检测模型

当窃电行为样本稀少或窃电手法不断更新、呈现新型特征时，获取足够数量且标注准确的“窃电”正样本变得极为困难，这使得依赖大量标注数据的监督学习模型面临严峻挑战，在此背景下，无监督或半监督学习的异常检测模型展现出独特价值和必要性，这类模型的核心思想是不依赖于已知的窃电样本，而是充分学习大量正常用户的用电行为模式，构建“正常”行为的基准模型或分布边界，进而识别出显著偏离该基准的异常个体，具体而言，孤立森林（Isolation Forest）算法随机分割数据来“隔离”样本，异常点因特征稀少而更容易被快速隔离，所以其路径长度较短，适用于检测用电量骤降、负荷曲线突变等明显偏离常态的行为；一类支持向量机（One-Class SVM）则在高维特征空间中寻找一个能包含大多数正常样本的最小超球面或边界，将位于边界之外的数据判定为异常，适合刻画复杂的行为轮廓；自编码器（Autoencoder）作为一种深度神经网络，编码-解码结构学习输入数据的低维表示，正常数据的重构误差较小，而窃电等异常行为因模式迥异导致重构误差显著增大,被识别出来，在实际部署中，这些模型主要利用未标注的用户数据或仅需少量确认的正常样本进行训练，即可对全量用户进行异常评分，比如某用户在夜间用电本应平稳的情况下突然出现周期性电量归零，孤立森林或自编码器可能迅速捕捉到这一异常模式，该方法的最大优势在于对未知窃电模式的适应性强，无需预先了解所有窃电类型，能够有效应对窃电手段的持续演化，是构建敏捷、鲁棒反窃电预警体系的关键技术路径。

（三）基于时序分析的深度学习模型

基于时序分析的深度学习模型在反窃电预警中具有独特优势,尤其适用于捕捉用户用电行为中蕴含的长期依赖关系、周期性规律以及复杂的动态变化模式，传统的统计方法难以充分建模用电数据中的非线性与时间序列特性，而长短期记忆网络（LSTM）和门控循环单元（GRU）等深度学习模型凭借其特有的门控机制，能够有效记忆和筛选长时间跨度内的关键信息，同时抑制噪声干扰，精准学习用户用电序列的内在规律，在实际应用中，首先利用大量正常用户的高频率用电数据（如每 15 分钟或每小时的用电量）对 LSTM 或 GRU 模型进行训练，使其掌握典型用电模式，如日周期（白天高峰、夜间低谷）、周周期（工作日与周末差异）以及节假日效应等，训练完成后，模型可用于滚动预测未来一段时间的用电量，持续监控实际用电量与模型预测值之间的偏差，一旦发现显著且持续的负向偏差（即实际用电远低于预测），则高度提示可能存在窃电行为，例如短接、分流或干扰电表等方式人为降低计量读数，为进一步提升判断准确性，可将预测误差与其他异常特征（如电压骤降、失压失流事件、频繁开盖记录等）进行融合分析，构建多维度的判别逻辑，这种基于深度学习的时序预测方法能实现对隐蔽性窃电行为的早期预警，

还具备较强的泛化能力，可适应不同用户类型和季节变化，为电力公司提供智能化、前瞻性的反窃电技术支撑。

（四）模型融合与预警生成

为了全面提升反窃电预警系统的准确性、全面性与鲁棒性，通常需要采用模型融合策略，将多种不同类型模型的优势进行集成，单一模型可能在特定场景下表现良好，但面对复杂多变的窃电行为时存在局限，所以加权平均、多数投票或元学习（Stacking）等融合技术，能够有效整合监督学习、无监督学习以及时序分析模型的输出结果，比如基于 XGBoost 等算法的监督分类模型可输出用户为窃电的概率值；孤立森林等无监督模型可提供异常得分，识别出偏离正常模式的行为;而 LSTM 等深度学习模型则预测用电量并计算实际与预测之间的偏差，捕捉潜在的时序异常，在融合阶段，可依据各模型在验证集上的性能（如 AUC、F1 值）动态分配权重，或采用硬投票/软投票机制，筛选出被多个模型共同判定为高风险的用户，随后，系统根据融合后的综合风险评分对所有用户进行排序，并结合历史稽查结果和业务需求设定动态阈值，生成高、中、低三级预警名单，电力稽查人员可据此优先对高风险用户开展现场核查，显著提升稽查效率与命中率，这种多模型协同的架构增强了系统对新型、隐蔽性窃电行为的识别能力,也提高了整体预警的稳定性与适应性，是构建智能化反窃电体系的核心环节。

三、挑战与对策

1.数据质量与完整性

部分老旧电表因设备老化或通信故障，存在数据采集不全、缺失或计量不准的问题，严重影响模型训练效果，为保障数据质量，应加强计量装置的日常运维与定期校验，及时更换故障设备，同时加快推进智能电表（AMI）的全面覆盖，提升数据采集的频率、精度与完整性，为反窃电模型提供稳定可靠的数据基础。

2.样本不平衡与标签获取难

窃电行为在整体用户中占比极低，导致正负样本严重失衡，且每起窃电需经现场稽查确认，标签获取周期长、成本高，为此可采用 SMOTE 等过采样技术平衡数据集，或使用代价敏感学习方法提升模型对少数类的识别能力，同时，结合无监督异常检测模型弥补标注数据不足,并建立窃电案例库，持续积累有效样本用于模型优化。

3.窃电手段的演化

窃电者可能研究反窃电机制,不断变换手法以规避检测，如采用间歇性窃电或新型干扰技术，对此，预警模型不能一成不变，必须建立动态更新机制，应定期引入新的行为特征与先进算法，结合现场稽查反馈的实际案例，持续训练和优化模型，提升其对新型窃电模式的识别能力，保持技术对抗的领先性。

4.隐私保护

在分析用户用电数据时，必须严格遵守《个人信息保护法》等相关法规，防范隐私泄露风险，应对涉及用户身份的信息进行脱敏处理，如加密用户编号、模糊化地址信息，确保分析过程仅基于用电行为特征,不关联可识别的个人身份，同时建立数据访问权限控制机制，保障数据使用合规、安全。

5.模型可解释性

深度学习等复杂模型虽预测精度高,但决策过程如同“黑箱”,影响稽查人员对预警结果的信任与采纳，为增强模型透明度，应引入 SHAP、LIME 等可解释性人工智能（XAI）技术，可视化展示影响预警的关键特征（如“日电量骤降 30%”或“频繁开盖”），帮助稽查人员理解判断依据，提升人机协同效率与决策可信度。

结论

基于大数据分析的反窃电预警模型，利用海量用电数据和先进的机器学习算法，能够有效克服传统方法的局限性，实现对窃电行为的精准、高效识别，构建融合监督、无监督及时序分析的多模型框架，电力公司可以显著提升反窃电工作的智能化水平，降低经济损失，维护电网安全与公平用电秩序，未来，随着物联网、边缘计算等技术的发展，反窃电模型将更加实时化、精细化，并与电网其他业务系统深度集成，构建全方位的智能用电管理生态。

[参考文献]

[1]张杰，蔺雪震，高燕增.基于大数据分析的供电所反窃电自动检验技术[J].电气技术与经济，2024(9)：113-115.

[2]邓丽娟.基于大数据技术的反窃电分析与仿真研究[J].电工材料，2022（6）：36-41.

[3]李端超，王松，黄太贵，等.基于大数据平台的电网线损与窃电预警分析关键技术[J].电力系统保护与控制，2018，46（5）：9.DOI：10.7667/PSPC170281.

[4]许明前，黄骁，陈富燕.基于大数据分析的反窃电实践[J].中国电力企业管理，2021（8）：1.

[5]张彤，穆秋羽.基于电力营销大数据技术的反窃电检查应用分析[J]. 2025（1）：85-87.