

# 基于大数据的反窃电智能识别模型及应用效果分析

白琳

内蒙古电力（集团）有限责任公司电力营销服务与运营管理分公司，内蒙古 呼和浩特 010020

DOI:10.61369/EPTSM.2025070004

**摘 要：** 电力在社会生产和人们生活中是不可或缺的能源之一，其能否稳定地供应对于社会经济的发展而言至关重要。但窃电行为一直困扰着电力行业的发展，它不仅会导致电力企业的经济效益受损，还极有可能引发电气火灾、线路过载等安全隐患，威胁着电网的安全运行和人民群众的生命财产安全。本文基于大数据的反窃电智能识别模型，阐述了模型的构建思路、核心架构与关键技术，且分析其在反窃电工作中的应用效果。

**关 键 词：** 大数据；反窃电；智能识别模型；应用效果

## Intelligent Recognition Model for Anti Electricity Theft Based on Big Data and Analysis of its Application Effect

Bai Lin

Inner Mongolia Electric Power (Group) Co., Ltd. Power Marketing Service and Operation Management Branch, Hohhot, Inner Mongolia 010020

**Abstract：** Electricity is one of the indispensable energy sources in social production and people's lives, and its stable supply is crucial for the development of the social economy. However, electricity theft has always plagued the development of the power industry. It not only damages the economic benefits of power companies, but also has a high risk of causing safety hazards such as electrical fires and line overloads, threatening the safe operation of the power grid and the safety of people's lives and property. This article is based on a big data based intelligent recognition model for anti electricity theft, elaborating on the construction ideas, core architecture, and key technologies of the model, and analyzing its application effect in anti electricity theft work.

**Keywords：** big data; anti electricity theft; intelligent recognition model; application effect

## 引言

传统的反窃电手段主要依赖于人工排查，但这种方式存在着诸多弊端。一方面人工排查需要耗费大量的人力、物力和时间，其效率低下，难以适应大规模、复杂的电力网络环境；另一方面因为人工排查过度依赖于工作人员的经验，主观性较强，极易容易出现漏查、误查等情况，最终导致窃电行为难以被及时地发现和查处。而现如今，随着大数据技术的迅猛发展，该技术在各个领域的应用都取得了显著的成效。若能将大数据技术引入反窃电工作之中，构建起反窃电智能识别模型，便能够实现对海量电力数据的深度挖掘和分析，助力供电企业及时地发现潜在的窃电行为，为反窃电工作提供科学、精准的决策支持。

## 一、基于大数据的反窃电智能识别模型

### （一）模型构建思路

基于大数据的反窃电智能识别模型的构建，核心在于充分地利用电力系统中产生的海量数据，经由对这些数据的分析和挖掘，从中找出与窃电行为相关的特征和规律，以此实现对窃电行为的精准识别<sup>[1]</sup>。但这一过程并非简单的数据堆砌与算法套用，

而是需要结合电力行业的业务特性与窃电行为的隐蔽性特点，构建起一套从数据到决策的完整逻辑闭环。

第一步，需要明确模型的目标是识别出存在窃电行为的用户或区域。为了实现该目标，需要提前收集与电力使用相关的各类数据，将这些数据将作为模型分析的基础，接着在目标设定阶段考虑不同场景下的窃电形态差异。例如居民用户可能会通过改装电表来实现窃电，工业用户则可能会采用绕越计量装置的方式偷

作者简介：白琳（1992.08-），女，内蒙古自治区包头市人，大学本科，中级工程师，研究方向：电力营销业务内控管理、用电检查、线损管理、数据治理。

电,那么模型就需具备对多类型窃电行为的覆盖能力。因此数据收集范围要兼顾多样性与针对性,当中既包括用户基础信息(如用户类型、用电地址、供电线路等),也应涵盖实时的用电数据(如分钟级用电量、电压电流曲线),还需纳入设备运行数据(如电表通信状态、故障告警记录)等等。

第二步,对于收集到的数据要进行处理和分析,从中提取出能够反映窃电行为的特征。通常这些特征可能包括用户的用电负荷曲线、用电量的波动情况、电表的运行状态等等。而特征地提取需建立在对窃电行为机理的深刻理解之上,如正常居民用户的用电负荷曲线一般呈现“早峰、午间低谷、晚峰”的周期性特征,窃电用户通过短时间内高频次断开计量回路,则会导致负荷曲线出现无规律的骤降。再如,商业用户在节假日的用电量会低于工作日,若某商铺在春节期间用电量异常偏低且与历史数据偏离度甚至超过阈值,提示该商铺可能存在着窃电嫌疑<sup>[2]</sup>。

第三步,基于提取出的特征选择合适的算法构建模型,随后通过模型对数据的学习和训练,使其具备识别窃电行为的能力。此时模型的训练需经历“样本标注—特征输入—参数迭代—效果验证”的循环过程。具体来说:先选取历史窃电案例与正常用电样本组成训练集,然后将标注后的样本特征输入算法进行学习,再通过调整算法参数来优化模型的性能,最终经由测试集来验证模型的识别准确率与召回率。

## (二) 模型核心架构

该模型的核心架构为数据层、特征层和模型层三个部分,这三层结构相互支撑、协同工作,共同实现了从原始数据到窃电识别结果的转化。首先是数据层,它是模型的基础,负责收集、存储和管理电力系统中的各类数据。因为这些数据的来源广泛,涉及到了用户的用电数据、电表的运行数据以及电网的相关数据。所以数据层需要具备强大的数据存储和管理能力,才能应对海量数据的处理需求。同时为了保证数据的质量和可用性,还需要对数据进行清洗和预处理操作,进而去除其中的噪声和异常值。其次为特征层,它的主要任务是从数据层提供的数据中提取出与窃电行为相关的特征,这也是模型构建过程中的关键环节,即特征的质量直接影响着模型的识别效果。实际在特征提取的过程中,特征层可被细分为“基础特征提取”与“高级特征构建”两个子模块,其中基础特征包括用户日均用电量、峰谷电比例、电压合格率等可直接从原始数据计算得到的指标,高级特征则需通过多维度的数据融合生成。此外特征层还需对特征进行标准化地处理,如将不同量级的用电量指标转换为[0,1]区间的归一化值,旨在避免因特征尺度的差异而影响模型训练效果的情况出现。最后则是模型层,该部分是整个智能识别模型的核心,主要负责利用特征层提取的特征进行模型的训练和预测。即在模型层中,需要选择合适的算法来构建模型。当前常见的算法有决策树、支持向量机、神经网络等<sup>[3]</sup>。而这些算法各有特点,适用于不同的场景和数据类型。结合实际应用来看,相关人员可能需要尝试多种算法,并通过比较它们的性能,再从中选择出最适合的算法来构建模型。

## (三) 模型关键技术

此模型的关键技术主要包括了数据融合技术、特征工程技术

和智能算法选择技术,基于上述三者共同构成模型的技术支撑体系,能够有效地保障模型的稳定性与识别精度。

### 1. 数据融合技术

电力系统中的数据来源多样,且不同来源的数据具有不同的格式和特点。而数据融合技术能够将这些不同来源、不同格式的数据进行整合和处理,消除不同数据之间的冗余和矛盾,有助于提高数据的一致性和准确性。例如将用户用电数据与电表GPS位置信息进行融合,即可识别出“一表多户”的窃电行为——当某块电表的注册地址为A小区,但其计量的用电量却与B小区某用户的实际用电轨迹高度吻合时,说明存在着电表被违规迁移的嫌疑。据此得知,经由数据融合能将用户的用电数据、电表的运行数据和电网的相关数据有机地结合起来,以此为特征提取和模型训练提供了更加全面、更丰富的信息。另外,数据融合还能实现跨系统数据的关联分析,比如将电力营销系统中的用户缴费记录与用电数据进行融合,若某用户长期拖欠电费且用电量远低于同类型用户,则可能存在着恶意窃电的倾向。

### 2. 特征工程技术

如前文所述,特征的质量直接影响着模型的性能。而特征工程技术分为特征选择和特征提取两个方面。一方面是特征选择,它是从众多的特征中筛选出对识别窃电行为最有价值的特征,去除当中无关或冗余的特征,以减少模型的复杂度,进而提高模型的运行效率。通常特征选择可采用基于树模型的特征重要性排序法——通过计算每个特征对模型决策的贡献度,然后保留贡献度排名前30%的特征,如“用电负荷波动率”“电表离线时长”等特征一般对窃电识别的贡献度较高,“用户姓氏”“电表生产厂家”等与窃电行为关联性较弱的特征则可被剔除<sup>[4]</sup>。另一方面是特征提取,它是通过对原始数据的变换和处理,生成新的、更能反映窃电行为本质的特征。即通过对用户用电量的归一化处理,消除不同用户之间用电量差异的影响,使特征更具有可比性,或者是计算“用电量环比增长率”,以此捕捉用户用电行为的短期异常变化。面对于工业用户,还可提取“功率因数异常频次”特征,因为正常工业用户的功率因数通常维持在0.9以上,但窃电用户可能通过接入感性负载改变了计量点的功率因数,导致功率因数会频繁的低于0.85。

### 3. 智能算法选择技术

由于不同的智能算法在处理不同类型的数据和问题时表现出不同的性能,因此需要根据实际的应用场景和数据特点,选择出合适的智能算法。实际在选择算法时,一定要考虑算法的准确率、效率、鲁棒性等因素:对于实时性要求高的场景(如台区线损率超标的即时排查),建议选择决策树、逻辑回归等运算速度快的算法;对于窃电手段复杂的工业用户识别,则可以选择深度学习算法,在借助多层神经网络捕捉数据中的深层特征;若面对的是数据样本不平衡的情况(如窃电样本仅占总样本的0.5%),便可以使用采用过采样、欠采样或集成学习等方法来改善算法性能。与此同时,还可以采用多种算法融合的方式,将不同算法的优势进行结合,进一步地提高模型的识别性能。举个例子,决策树算法的可解释性强和神经网络算法的非线性拟合能力强的特点

相结合，构建起混合模型——先用神经网络对高维特征进行降维处理，再将降维后的特征输入决策树生成识别规则。此举既保证了模型对复杂模式的捕捉能力，还能提高了识别结果的可解释性，更方便电力稽查人员理解模型判定的依据<sup>[5]</sup>。

## 二、基于大数据的反窃电智能识别模型应用效果分析

### （一）提升反窃电工作效率

在传统的反窃电工作过程中，主要依靠的是人工排查的方式，该方式需要工作人员逐个用户、逐个区域的进行检查，工作量巨大且效率低下。基于大数据的反窃电智能识别模型则能够实现海量电力数据的自动分析和处理，进而快速地识别出存在着窃电嫌疑的用户或区域，在极大程度上减少了人工排查的范围和工作量。而模型在短时间内对大量用户的数据进行分析后，会将存在异常用电行为的用户筛选出来，工作人员只需针对这些筛选出的目标进行重点检查即可。如此一来，便避免了盲目地排查，提高了反窃电工作的效率。

### （二）提高窃电行为识别精准度

由于传统反窃电手段过度地依赖于人工经验，非常容易受到工作人员主观因素的影响，导致窃电行为的识别精准度不高，其中存在着漏查和误查的情况。在以大数据为基础建立的反窃电智能识别模型之中，通过对海量数据的分析和挖掘，能够发现一些人工难以察觉的细微特征和规律，使窃电行为的识别精准度得到了提高。因为模型可以学习大量的历史窃电案例数据，并总结出窃电行为的特征模式，然后将这些模式应用到新的数据中，就能对用户的用电行为进行判断。哪怕是一些隐蔽性较强的窃电行为，如通过修改电表参数、私拉乱接线路等方式进行窃电的行为，模型依然能够通过分析电表的运行数据和用户的用电曲线等信息，及时地发现异常情况，提高了识别的准确性。

### （三）降低反窃电工作成本

传统的反窃电工作往往需要投入大量的人力、物力和财力，原因是人工排查需要配备足够的工作人员，还需要使用各类检测设备和工具，而这些都会产生较高的成本。但基于大数据的反窃

电智能识别模型能够在一定程度上减少对于人工的依赖，间接地降低了反窃电工作的成本。此方面最为关键的就是模型的应用减少了人工排查的工作量，大幅度地减少了人力成本的投入。加之模型能够精准地识别出窃电嫌疑目标，以此避免了不必要的设备使用和资源浪费，所以降低了物力和财力成本<sup>[6]</sup>。举个例子，经模型筛选出的可疑用户数量有所减少，工作人员便可以集中使用检测设备对这些用户进行检查，这样便提高了设备的使用效率，从而降低了设备的损耗和维护成本。

### （四）保障电力系统安全稳定运行

窃电行为不但会造成经济方面的损失，还可能会对电力系统的安全稳定运行造成一定的威胁。比如，窃电行为可能会导致线路过载、电压不稳定等问题出现，随即影响到其他用户的正常用电，甚至还会引发电气火灾等安全事故。借助反窃电智能识别模型，便能够及时地发现窃电行为，以便电力企业采取相应的措施对其进行处理，有效地保障了电力系统的安全稳定运行。展开而言：基于模型的实时监测和分析，电力企业可以在窃电行为造成严重后果之前及时地介入，提前制止窃电行为，并修复被破坏的电力设施。同时模型还可以为电力企业提供关于窃电行为的统计分析和趋势预测，帮助企业制定出更加有效的防范措施，直接从源头上减少窃电行为的发生，促使电力系统的安全性和可靠性得到提升。

## 三、结语

结合了大数据的反窃电智能识别模型，为电力行业的反窃电工作提供了一种全新的、高效的技术手段。该模型通过对海量电力数据的分析和挖掘，能够精准地识别窃电行为，进而有效地提升了反窃电工作的效率和质量，并降低了工作成本，同时也为保障电力系统的安全稳定运行发挥了重要的作用。然而，该模型在实际应用之中仍存在一些不足之处。例如模型对数据的质量和完整性要求较高，若数据存在着缺失或错误，便可能会影响到模型的识别效果。此外随着窃电手段的不断翻新，模型需要持续地进行更新和优化，以适应新的挑战。

## 参考文献

- [1]张靖.基于大数据的用电稽查研究[D].北京市:华北电力大学,2019.
- [2]刘静.江门电网窃电辨识与反窃电应用技术研究[D].广东省:华南理工大学,2016.
- [3]黄铁柱.大数据技术在反窃电检查中的应用研究[J].IT经理世界,2024,(10):133-135.
- [4]卢娜,王辉,刘浏.基于大数据分析的精准反窃电实践探究[J].黑龙江科学,2023,14(18):155-157+161.
- [5]黄春旭.基于用电信息采集的智能反窃电识别技术分析[J].电子技术(上海),2024,53(04):258-259.
- [6]河南光石智能工程有限公司.基于用电信息采集大数据分析的反窃电智能分析识别设备:CN216209408U[P/OL].2022-04-05[2025-08-18].<https://www.cqvip.com/doc/pat-ent/1970590116>.