

基于CNN-SVM模型的电网窃电风险自动化预警算法

张春梅¹, 袁杰生¹, 许兴雀¹, 黄彬系², 王曦¹

(1. 广东电网有限责任公司中山供电局, 广东 中山 528400;

2. 广东电力信息科技有限公司, 广东 广州 510030)

摘要: 针对电网窃电风险预警时效性差、精度低的问题, 设计基于CNN-SVM模型的电网窃电风险自动化预警算法。分别从日电压均值、功率因数、电流信息等方面采集用户用电数据, 利用最小-最大标准化方法预处理数据; 构建窃电风险预警参数集合, 确定暂态因子权值, 将电压扰动、负载波峰作为特征指标, 提取窃电风险边界特征; 建立SVM分类器模型, 获得最优分类平面和决策函数; 通过卷积神经网络改进SVM分类器, 构建CNN-SVM模型, 提高模型的学习能力; 经过模型逐层学习, 输出最终特征向量, 即窃电风险等级, 结合该等级实现自动化预警。实验结果表明, CNN-SVM模型的收敛效果好, 训练耗时短, 能够准确识别出窃电用户, 应用于电网窃电风险自动化预警中的预警精度高, 响应速度快。

关键词: 支持向量机; 卷积神经网络; 窃电风险; 自动化预警; 决策函数

中图分类号: TP183 文献标识码: A 文章编号: 1003-7241(2025)04-0038-06

Automatic Early Warning Algorithm of Power Theft Risk Based on CNN-SVM Model

ZHANG Chun-mei¹, YUAN Jie-sheng¹, XU Xing-que¹, HUANG Bin-xi², WANG Xi¹

(1. Zhongshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Zhongshan 528400 China;

2. Guangdong Electric Power Information Technology Co., Ltd., Guangzhou 510030 China)

Abstract: Aiming at the problems of poor timeliness and low accuracy of power theft risk early warning, an automatic power theft risk early warning algorithm based on CNN SVM model is designed. It collects user power consumption data from aspects such as daily voltage average, power factor, and current information, and preprocesses the data using the minimum maximum standardization method, establishes a set of early warning parameters for power theft risk, determines the weight of transient factors, and extracts the boundary characteristics of power theft risk using voltage disturbances and load peaks as characteristic indicators, establishes a SVM classifier model to obtain the optimal classification plane and decision function. Using Convolutional Neural Networks to improve SVM classifiers, construct CNN-SVM models, and improve the learning ability of the models. After learning the model layer by layer, the final feature vector, namely, the risk level of electricity theft, is output, and combined with this level to achieve automatic early warning. The experimental results show that the CNN-SVM model has good convergence effect, short training time, and can accurately identify power theft users. The early warning accuracy and response speed applies to the automatic early warning of power theft risk in power grids are high.

Keywords: support vector machine; convolutional neural network; electricity theft risk; automatic early warning; decision function

0 引言

近年来, 电力建设脚步加快, 但部分地区还存在供应薄弱、窃电漏电等现象。随着用户用电需求的增加, 一些非法用户在利益驱动下通过各类方式实施窃电行为, 不仅导致电能损失, 提升线损率, 而且还干扰了正常用电秩序, 给企业带来经济损失。因此, 如何保证电力系统的稳定运行, 及时地发现和处

理电力系统亟待解决的重要课题。随着人工智能技术的飞速发展, 许多学者都将人工智能技术用于电力系统窃电诊断与预警当中, 开展了很多人工智能算法应用于反窃电方面的工作。

例如, 国内学者蔡嘉辉^[1]等人利用随机森林网络算法监测窃电行为。采用神经网络提取大规模用电数据的窃电特征, 结合这些特征建立随机森林训练分类器, 使用网格搜索法选择最佳参数, 检测出用户是否存在窃电行为。唐东来^[2]等人研究一种基于时空关联矩阵的窃电预警算法。针对采集的海量用电数据进行清洗, 使用线性插值法填补缺失值; 设计窃电分析算法, 结合线损波动、

*基金项目: 中山供电局2021年综合城个性化数据应用建设(全业务合规管理监控应用)项目(032000HK42210004)

收稿日期: 2023-11-22

电流差异特点等指标进行时空关联分析,判断出用户窃电行为。国外一些学者的研究方向也大多集中在线损信息监测方面,通过抄表功能,获取用户电量信息,结合线损率等指标锁定嫌疑用户。

随着检测技术的发展和长时间的经验累积,自动化预警已经日趋完善。为此,本文利用卷积神经网络-支持向量机模型(convolutional neural network-support vector machine, CNN-SVM)方法实现电网窃电风险的自动化预警。支持向量机(support vector machine, SVM)属于一种监督学习方法,是对数据进行分类的线性分类器^[3],最大超平面是其决策边界,影响着分类效果。通过卷积神经网络(convolutional neural network, CNN)改进SVM模型,可提高算法的实时性,更有利于特征学习,提高预警结果精度。所提方法不但可以尽早发现用电异常情况,还能有效维护电力企业的利益,对净化电力市场和维护用电秩序意义重大^[4]。

1 电网用电数据采集与处理

1.1 用电数据采集

用户的用电行为包括很多种,而用电指标能够从各个方面反映用电的内在规律。为全面采集用电数据,需要综合分析多个指标,实现对用户窃电风险的全方面评价。

当窃电行为发生时,用电模式必然会发生异常,且用电数据通常不是孤立的,在相同窃电场景中会包含多种异常^[5]。如果只采集单方面数据,有可能会提高虚警率。因此,本文分别从下述几方面采集用电数据:

(1) 日电压均值:合法用户的电压基本没有太大波动,当发生电压异常时,表明用户存在窃电的可能;

(2) 功率因数:在负载特征作用下,功率因数浮动较小,数值比较稳定,通常不会发生突变现象;

(3) 电流信息:电流随用户负载的变化也会出现无规律改变,正常情况下,用户的相线与零线的电流量基本相同;

(4) 日冻结电量:如果某个用户的用电行为可疑,则该用户的电量信息与其他用户相比是异常的,结合此类用户的申报容量,能够判断出是否存在异常;

(5) 电量不平衡率^[6]:针对某一用户,持续采集电流信息,根据电压均值与功率因数,得出理论电能增量 W_1 ,并将该值与此时段内电能增量对比,观察二者是否不平衡。计算公式如下:

$$\eta = \frac{|W_1 - W_2|}{\max\{W_1, W_2\}} \quad (1)$$

式中, η 为电量不平衡率, W_2 是实测值。

综合分析抄表系统的容量情况,在每次采集过程中只对同一批次中少数用户进行采集,确保每分钟一次的

采集频率,且对于其他用户通过交叉时段采集方式完成数据采集工作。

1.2 数据预处理

随着电力用户的数量越来越多,且受采集设备通信情况影响,采集过程中具有很多不确定因素,容易受到噪声干扰,出现数值缺失现象。这会严重影响对窃电用户数据的分析结果,导致预警精度下降。为此,有必要对初始数据做预处理,提高数据质量。

本文利用最小-最大标准化来处理初始数据^[7],处理后的数据将会限制在 $[0, 1]$ 范围内,具体处理过程如下用公式描述:

$$v_i = \eta \frac{v_i - v_{\min}}{v_{\max} - v_{\min}} \quad (2)$$

式中, v_i 表示初始数据, v_i' 是处理后的数据, v_{\min} 、 v_{\max} 分别是对应属性的极大值与极小值。

2 基于 CNN-SVM 模型的电网窃电风险自动化预警

在实现了电网用电数据采集和处理的基础上,构建 CNN-SVM 模型实现电网窃电风险自动化预警。首先提取窃电风险边界特征,然后利用 CNN 模型改进优化 SVM 模型,通过 CNN 模型训练电网用电数据,利用卷积层降低数据维度,得到准确的电网用电数据训练结果,再将处理结果输入到 SVM 分类器中,输出最终窃电风险等级分类结果,实现窃电风险自动预警。

2.1 窃电风险边界特征提取

为实现窃电风险的自动化预警,需要提取风险特征向量。将采集的电流、电压、功率因数等信息作为基础,建立电网窃电风险预警的参数集合^[8]。通过下述公式描述电网安全运行时的暂态因子权值:

$$G_{k+1} = v_i' P_{k+1|k} - \sqrt{P_{k+1|k}} \quad (3)$$

式中, $P_{k+1|k}$ 和 $P_{k+1|k}$ 分别代表电网约束指标和风险评估指标, k 属于扰动参数。

在参数扰动情况下,电网负载信息^[9]计算公式如下:

$$H_{k+1} = \frac{P_{k+1|k}}{G_{k+1}} + \sum_{k=1}^{\infty} P_{k+1|k} \quad (4)$$

式中, H_{k+1} 为电网负载量, $P_{k+1|k}$ 是扰动因子矢量。

在自然换流情况下,将电压作为主要窃电分析指标,利用下述公式计算窃电时输出和输入电压的比值:

$$V(t) = H_{k+1} \left[\frac{j(t)}{V_x(t)} - \frac{\pi}{k^2 n^2} \right] \quad (5)$$

式中, $V_x(t)$ 代表电压变换函数, n 是风险分析的测度集合, $j(t)$ 属于固定电流函数。则电压扰动表示为:

$$V' = V(t) \int_{-\infty}^{\infty} \sin(\omega t) - \frac{\theta}{\omega t} \quad (6)$$

式中, ωt 是描述电压稳定性, θ 描述特征点向量。

探究电网窃电风险的向量位数, 通过修正函数 $\rho(k)$ 完成位数统计, 获得下述非线性调整因子 φ_k :

$$\varphi_k = V' \int_{k=1}^n \rho(k) \omega t - \sqrt{2\pi} n + \frac{r_k}{R_k} \quad (7)$$

式中, r_k 代表风险维数, R_k 是修正因子, 不同时间段内 r_k 和 R_k 值也不同。

将负载波峰作为特征指标, 利用下述公式提取出风险评估的边界特征^[10]:

$$R_{k+1} = V_M (\hat{q}_{k+1} - \hat{Q}_{k+1} + \varphi_k) + \sum_{k=1}^n \hat{r}_{k+1} \quad (8)$$

式中, V_M 为边界集合, \hat{q}_{k+1} 和 \hat{Q}_{k+1} 分别为负载变化和电网安全指标。

利用上述方法, 获得风险边界的特征矢量, 结合该风险特征, 通过 CNN-SVM 模型对用户数据进行分类, 通过迭代训练得到风险输出结果, 对输出结果进行分类, 即可有效识别出风险用户, 实现自动化预警。

2.2 基于 CNN-SVM 模型的自动化预警实现

SVM 属于线性模型分类器, 通过非线性处理, 将输入空间变换为特征空间, 获得最佳线性平面。一般情况下维数决定了分类效果, 为方便计算, 通常使用核函数提高分类精度。

假设线性可分样本表示为 $(x_a, y_a), a=1, 2, \dots, N$, 其中 x_a 是样本矩阵, $y_a \in \{-1, 1\}$ 是样本标签。判别函数表示为 $h(x) = ax + b$, 则分类面的表达式为:

$$\alpha x + b = 0 \quad (9)$$

针对上述公式做归一化处理, 通过等比方法调节矢量系数 α 和 b , 则分类曲面寻优问题变换为:

$$\min(\alpha) = \frac{1}{2} \|\alpha\|^2 = \frac{1}{2} (\alpha^2) \quad (10)$$

$$s.t. \ y_a [(ax_a) + b] - 1 \geq 0$$

因此, 能够获得最佳决策函数:

$$f(x) = \text{sgn} \left(\sum_{a=1}^n \delta_a y_a (x_a + B) \right) \quad (11)$$

式中, $\text{sgn}()$ 代表符号函数, δ_a 属于拉格朗日乘数, B 代表分类阈值。

利用核函数 $K(X_p, X_q)$ 描述向量内积, 则优化问题的表达式如下:

$$\min(\alpha) = \frac{1}{2} \sum_{p,q=1}^a \delta_p y_p (X_p, X_q) - \sum_{p,q=1}^a \delta_a \quad (12)$$

$$s.t. \ \delta_a \geq 0, \sum_{p,q=1}^a y_p \delta_p \geq 0$$

与其对应的决策函数如下:

$$f(x) = \text{sgn} \left(\sum_{a=1}^a \delta_a K(X_p, X_q) + B \right) \quad (13)$$

上述 SVM 模型虽然能够得到很好的数据分类结果, 可在用户样本中确定窃电风险用户。但是维数灾难问题突出, 计算量非常大, 难以实现实时预警。

针对上述经典 SVM 模型, 使用卷积神经网络对其改进优化, 设计改进的 CNN-SVM 模型, 将提取到的特征作为 CNN 模型迭代训练的依据, 在其他参数共同作用下, 获得新的训练特征, 再将该特征输入到 SVM 分类器中, 得到最终预警结果。

CNN 是一种人工神经网络, 利用卷积运算达到数据非线性变化的目的。在 CNN 结构中, 嵌入层作为一个矩阵, 能够实现数据状态向量的排序; 卷积层通过运算可获得很多种特征映射, 再传输到池化层减少数据维度; 池化层最后与全连接层相连, 输出训练结果。改进后的模型结构如图 1 所示。

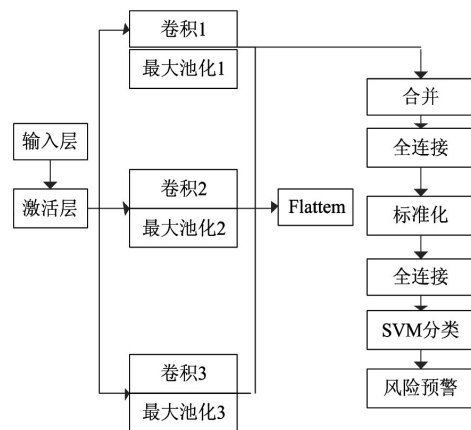


图 1 CNN-SVM 模型结构示意图

由图 1 可知, 模型的首层为输入层, 其次是预处理层, 第三层为并行的卷积神经网络模型, 该模型为双卷积形式。为了更好地完成特征学习, 可以通过改变窗口大小来提高模型学习性能, 且将 CNN 模型和 Flatten 层相连, 有效降低数据维度。

假设窗口长度表示为 L , 当卷积核 D 完成运算后, 输出的运算结果表示为:

$$O_e = \frac{f(x) R_{k+1} (w T_e + d)}{\sigma} \quad (14)$$

式中, T_e 代表输入矩阵, w 是权重矩阵, d' 属于偏置因子, σ 表示激活函数。

在上述构建的模型中, 激活函数利用 Tanh 函数, 且模型具备两个全连接层^[11], 主要作用是对特征做加权处理, 输出特征向量:

$$U' = \phi \frac{O_e (w + u)}{d'} \quad (15)$$

式中, ϕ 和 u 分别描述上、下层的输出向量。

将采集到的用户用电数据在窗口中转换为一定长度的向量 U^1 , 结合提取的风险边界特征, 利用 CNN 模型进行处理, 将处理结果输入到 SVM 分类器中, 输出最终窃电风险分类结果, 针对窃电风险等级较高的用户进行实时自动预警, 电力企业根据预警结果采取相应措施, 避免窃电事件发生, 保证电网运行安全^[12]。具体预警流程如图 2 所示。

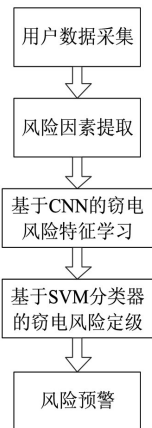


图2 窃电风险自动预警流程图

3 实验过程与结果分析

3.1 实验环境及实验数据

为证明 CNN-SVM 模型技术在窃电风险自动预警方面的实际应用效果, 设置如下实验。目标电网的绕组电压是 150 kV, 调频为 15 kHz, 额定电流为 200 nA。实验数据集合来自国网电网公司 2021 年 1 月-12 月份内 52 425 个用户的用电数据。实验平台配置如图 3 所示。

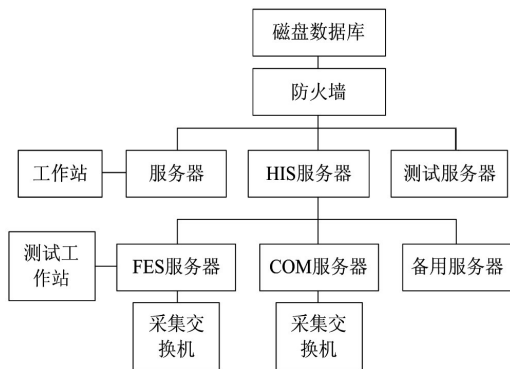


图3 实验平台结构配置图

3.2 CNN-SVM 模型性能验证

首先分析 CNN-SVM 模型的性能, 选取用电数据样本数量为 16 000 个, 验证 CNN-SVM 模型的训练耗时。将本方法测试结果和随机森林算法、时空关联矩阵算法结果进行对比, 如图 4 所示。

分析图 4 可知, 随着数据样本数量的增多, 所有方法的训练耗时均有所增加。当样本数量达到 16 000 个时,

随机森林算法的训练耗时为 11 s, 时空关联算法的训练耗时为 8.5 s, 本文方法的训练耗时为 3.2 s, 且本文方法的训练耗时随样本增加而增长的幅度较小。

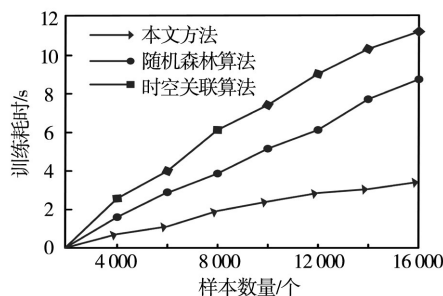


图4 不同方法数据样本训练耗时测试结果

在此基础上, 测试三种方法对数据样本训练过程的迭代收敛性能, 结果如图 5 所示。

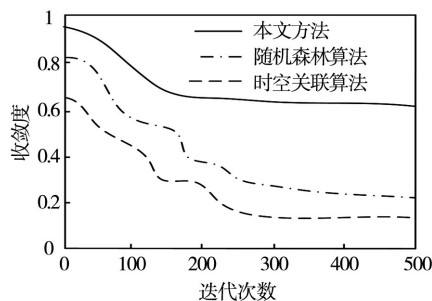


图5 不同方法的迭代收敛性能对比

分析图 5 可知, 本文构建的 CNN-SVM 模型在迭代次数为 130 次时就达到了收敛, 收敛度为 0.68, 随机森林算法在迭代次数为 220 次时达到了收敛, 收敛度为 0.26, 时空关联算法在迭代次数为 210 次时达到了收敛, 收敛度为 0.16。由此可见, 本文所构建的 CNN-SVM 模型训练用电数据样本的收敛性能更好, 能够较快完成对用电数据样本特征样本的训练, 有效提升数据样本分类识别能力。

3.3 电网窃电风险自动化预警实验结果

实验选取的数据集合包括 52 425 个用户的用电数据, 其中窃电用户有 60 个。用户分为工业用电用户、商用用电用户、居民用电用户 3 类。分别采用本文方法、随机森林算法、时空关联矩阵算法对 52 425 个用户的用电数据进行窃电风险自动化预警, 验证这三种方法能够准确预警 60 个窃电用户的风险, 得到实验对比结果如表 1 所示。

表1 电网窃电风险自动化预警结果

方法	正常数据	窃电风险预警结果
随机森林算法	52 351	46
时空关联算法	52 356	51
本文方法	52 365	60

根据表 1 可知, 本方法能够准确预警窃电用户, 预警结果与实际窃电用户个数一致, 而随机森林和时空关联算法无法准确预测出所有的窃电用户。由此可见, 本方法

能够有效识别出窃电用户,准确预警电网窃电风险。

分析CNN-SVM模型对于非法用户用电数据的分类预警能力,将 F_1 值和AUC(曲线下面积)作为评价指标。 F_1 值表示准确率与召回率的平均水平,计算公式如下:

$$F_1 = \frac{2PR'}{P+R'} \quad (16)$$

式中, P' 与 R' 表示准确率与召回率。由此可知, F_1 值越大,说明算法的分类预警能力越强。将表1数据代入所测试模型,将本方法测试结果和随机森林算法、时空关联矩阵算法结果进行对比,如图6所示。

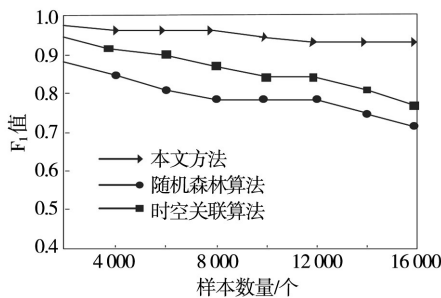


图6 不同方法数据样本分类精度测试结果

由图6可以看出,随着数据样本数量的增多,所有方法的 F_1 值均有所下降。其中随机森林和时空关联算法的曲线趋势相同,下降幅度较大,本方法没有出现明显下降趋势,始终保持在0.9以上,且当样本数量为12 000时,曲线几乎无波动,当样本继续增加时,基本不会继续下降。由此可见,电网窃电风险自动化预警的数据样本分类精度较好。

AUC表示坐标轴中 $x=y$ 曲线和坐标轴形成的面积,该面积通常小于1。因此,AUC取值在0.5-1之间,且越接近1说明预警效果越好,当低于0.5时表明算法无应用价值。指标公式如下:

$$AUC = \frac{\sum_{i=1}^{n} rank_i - \frac{\alpha(1+\alpha)}{2}}{\alpha \times \beta} \quad (17)$$

式中, $rank_i$ 代表样本 i 的排序值, α 与 β 分别为正、负样本总数。上述三种方法的实验效果如图7所示。

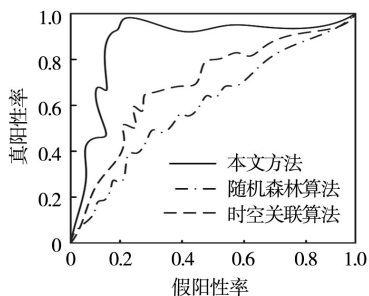


图7 不同方法AUC曲线图

观察图7得出,本方法的曲线更加靠近上方,说明该

曲线与 $x=y$ 曲线组成的面积最大,预警效果最好。这是因为,CNN-SVM模型训练用电数据样本的收敛性能更强,数据分类能力较好,根据提取到的窃电数据特征,CNN-SVM模型分类器可以有效识别窃电用户,提高预警效果。

一般的预警算法会占用较多的系统内存,且随着用户数量的增多,预警响应时间也会逐渐增加,难以实现实时预警。为此,本文测试了算法的并发能力。分析算法能否快速对用户用电数据做出响应。测试结果如图8所示。

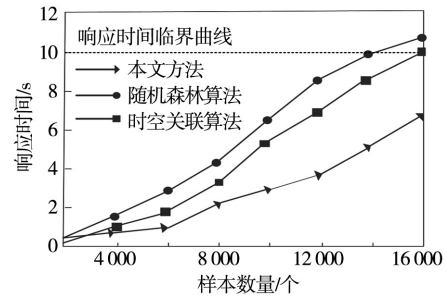


图8 不同方法的预警响应时间测试结果图

由图8能够看出,随着并发用户数量的增多,三种方法的响应时间都随之增加,但本方法始终没有超过响应时间临界曲线,说明该方法实时预警效果更好。这是因为利用卷积神经网络对SVM算法进行改进,得到的CNN-SVM模型有效提高了学习速率,减少了电网窃电风险自动化预警时间。

4 结束语

窃电行为长期危害着电力企业的利益,对社会经济发展的影响也较大。本文提出了基于CNN-SVM模型的电网窃电风险自动化预警算法。通过数据挖掘的方式对用电信息进行分析,采用数据预处理技术提高减少干扰,提取窃电风险数据的特征,再使用卷积神经网络CNN-SVM模型,改善学习能力,提高预警精度。实验结果表明,用电数据中包含很多信息,挖掘这些隐藏信息可以满足窃电检测的需求,进而得到精准的预警结果。

参考文献:

- [1] 蔡嘉辉,王琨,董康,等.基于DenseNet和随机森林的电力用户窃电检测[J].计算机应用,2021,41(S1):75-80.
- [2] 唐冬来,刘友波,熊智临,等.基于时空关联矩阵的配电台区反窃电预警方法[J].电力系统自动化,2020,44(19):168-176.
- [3] 陈文瑛,龙跃,傅宏,等.客户侧窃电态势感知及智能预警关键技术的研究[J].电子技术应用,2021,47(12):69-73.
- [4] 徐晓虹,郑昀.基于大数据的配电设备运行状态监控与故障预警系统设计[J].能源与环保,2021,43(12):207-211.
- [5] 殷涛,薛阳,杨艺宁,等.基于向量自回归模型的高损线路窃

(下转第88页)