基于机器学习的电力互联网攻击信息识别方法研究

杨东宁,张志生,张万辞,刘 鑫
(云南电网有限责任公司信息中心 云南 昆明 650000)

摘要: 为解决传统攻击信息识别方法存在识别误差大的问题,提出基于机器学习的电力互联网攻击信息识别方法。依据互联网攻击信息,构建互联网攻击信息模型,分析基于机器学习的电力互联网攻击信息识别原理,结合哈希定值保障相同攻击信息会分配到同一线程之中,避免噪声产生的偏差,实现电力互联网攻击信息的实时无损处理。构建脆弱性邻接矩阵,并对脆弱性进行定量评估,完成电力互联网攻击信息优化识别方案设计。实验结果表明,该方法识别精度最高可达到98%,能够有效降低电力互联网网络攻击风险,保障网络安全稳定运行。

关键词: 机器学习; 电力互联网; 攻击信息; 识别
中图分类号: TN49 **文献标识码:** A **文章编号:** 1674-6236(2020)17-0066-04
DOI: 10.14022/j.issn1674-6236.2020.17.015

Research on the method of power Internet attack information recognition based on machine learning

YANG Dong-ning, ZHANG Zhi-sheng, ZHANG Wan-ci, LIU Xin
(*Information Center of Yunnan Power Grid Co., Ltd., Kunming* 650000, *China*)

Abstract: In order to solve the problem of large error in traditional attack information recognition methods, a machine learning based attack information recognition method for power Internet is proposed. According to the Internet attack information, the Internet attack information model is constructed, the principle of power Internet attack information recognition based on machine learning is analyzed, and the same attack information will be allocated to the same thread with hash fixed value guarantee, so as to avoid the deviation caused by noise and realize the real-time lossless processing of power Internet attack information. The vulnerability adjacency matrix is constructed, and the vulnerability is quantitatively evaluated to complete the design of optimal identification scheme of power Internet attack information. The experimental results show that the recognition accuracy of this method is up to 98%, which can effectively reduce the risk of power Internet network attack and ensure the safe and stable operation of the network.

Key words: machine learning; power Internet; attack information; recognition

为了满足用户需求,电力互联网不管是对网络资源进行管理,还是对网络进行改造升级,都需对网络中各种应用有所了解[1]。电力系统信息化进程加快,也为电力互联网带来不可忽视的网络安全问题,在该背景下,对电力互联网攻击信息识别可以有效解决上述问题[2]。

文献[3]提出基于综合权重法的电力互联网攻击信息识别方法,利用综合权重理论,将电力互联网攻击图生成过程划分为多个子区域,每个区域都可实现两个电力互联网信息子网中攻击信息图的生成,依据该信息聚合全部子攻击图,完成电力互联网攻击信息识别。虽然该方法扩展性较强,但耗费时间较长。文献[4]提出基于TOPSIS算法的攻击信息识别方法,利用TOPSIS算法构建攻击信息图,生成攻击识别优化方案,实现电力互联网对攻击信息的识

收稿日期: 2019-11-22 **稿件编号:** 201911162
作者简介: 杨东宁(1986—),男,云南昆明人,硕士,工程师。研究方向:电网信息化。

别。该方法虽然能够缩短攻击信息识别时间,但存在攻击信息识别效果差的问题。针对以上问题,提出基于机器学习的电力互联网攻击信息识别方法。

1　攻击信息识别原理

构建电力互联网攻击信息建模后,对电力互联网空间中的攻击行为进行模型化分析,并以此指导互联网攻击的防御。由于互联网模型不能描述攻击持久性,也就无法反映攻击场景特征,因此,将互联网攻击信息模型进行扩展,如图1所示。

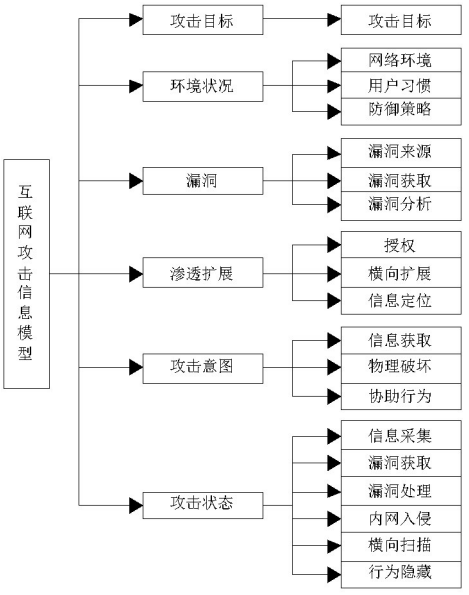

图1　互联网攻击信息模型

采用机器学习方法模拟人类学习行为,以此获取新的知识[5]。环境向系统学习部分提供某些信息,系统利用这些信息修改知识库,以此增进攻击信息识别任务的完成能力,同时将获取的信息反馈给学习部分[6]。环境所提供的信息是影响机器学习的主要因素,如果信息质量较高,那么与一般原则相比,机器学习部分信息是容易处理的。如果向学习系统提供具体信息,那么学习系统需要在获取足够信息之后,删除多余信息,进而总结扩展,形成指导动作原则[7]。通过机器学习识别方法,可留下正确规则,剔除不正确规则,并从数据库中删除,在识别过程中,需遵照以下4个原则,分别是表达能力强、方便推理、容易修改知识库、方便扩展[8]。以此为依据,识别电力互联网攻击信息。按照学习形式,将机器学习分为3种,分别是监督学习、无监督学习和半监督学习,机器学习是从环境中不断采集新的信息,并更新已经学会的知识体系,可以不断提高电力互联网攻击信息的识别效率[9-11]。

主要识别流程为:

1)从电力互联网攻击信息中采集相关信息,并剔除冗余信息。

2)向知识库中存放正确、有价值的知识信息;

3)根据采集到的有效信息进行归纳处理;

4)依据知识库中已经存在的规律,解决电力互联网攻击问题;

5)从电力互联网中采集有效信息并传递给学习模块,以此实现攻击信息识别[12]。

2　攻击信息实时无损处理

基于监督机器学习进行有效线程规划,结合哈希定值保障相同攻击信息会分配到同一线程之中进行处理,避免产生偏差[13-14]。利用线程池调度管理,避免大量攻击信息间的相互排斥。无损处理技术结构如图2所示。

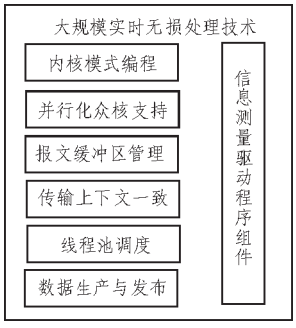

图2　电攻击信息实时无损处理

对电力互联网攻击信息实施无损处理能够有效发挥并行计算能力,使用被动线程调度方式,动态调整用于分析处理的线程数量,以确保实现电力互联网攻击信息的实时处理[15]。

3　攻击信息优化识别

3.1　组建脆弱性邻接矩阵

假设电力互联网中的安全漏洞为潜在攻击信息,脆弱性邻接矩阵中元素 $q_{i,j}$ 表示从电力互联网攻击节点 i 开始通过进一步攻击成功接入到节点 j 的可能性,利用如下公式计算 $q_{i,j}$:

$$q_{i,j}=\begin{cases}\alpha_{acc,j} & i=1\\ \beta_{i,j} & i\neq 1\end{cases} \tag{1}$$

式(1)中:当 $i=1$ 或 $j=1$ 时,表示该处为攻击节

点；$\alpha_{acc,j}$ 表示选取节点为并行攻击接入点出现的可能性；$\beta_{i,j}$ 表示两个节点间出现攻击信息的可能性[16]。

利用机器学习方法计算电力互联网攻击长度的 N 步脆弱性矩阵，再将每个矩阵所对应位置上的元素进行求和处理，即可计算出选取节点为攻击节点的可能性，即每个网络节点脆弱性量化值，利用如下公式表示：

$$W_{i,j}=\sum_{n=1}^{N}q_{i,j} \tag{2}$$

式(2)中：$W_{i,j}$ 表示网络节点脆弱性量化值。

假设，各个接入点在接入后对电力互联网所产生的影响均不相同，通过式(3)计算电力互联网中任意安全漏洞被攻击时可接入点的可能性指标：

$$E_{s,i}=\begin{cases}\dfrac{T_{v,i}T_{a,i}}{\sum\limits_{j\in acc}T_{v,j}T_{a,j}} & i\in S_{acc}\\ 0 & i\notin S_{acc}\end{cases} \tag{3}$$

式(3)中：$T_{v,i}$ 表示安全漏洞价值；$T_{a,i}$ 表示电力互联网安全漏洞接入难度；$T_{v,j}$ 表示电力互联网安全漏洞接入的速度；$T_{a,j}$ 表示电力互联网安全漏洞风险；S_{acc} 表示电力互联网中有可能被选取攻击点的安全漏洞集合。

假设，$Q_{v,i}$ 表示电力互联网安全漏洞被攻击成功后对网络造成的影响，利用如下公式表示：

$$Q_{v,i}=A_{acc,i}\cdot\theta_i \tag{4}$$

式(4)中：$A_{acc,i}$ 表示安全漏洞所在电力设备资源吸引力；θ_i 表示安全漏洞影响因子。

3.2 攻击信息优化识别方案设计

在进行电力互联网攻击信息优化识别过程中，以上述组建的脆弱性邻接矩阵为依据，获取攻击信息，计算每个攻击路径被成功攻击概率，得到并行攻击度量指标，以此为依据识别电力互联网攻击信息。电力互联网攻击信息优化识别流程如图3所示。

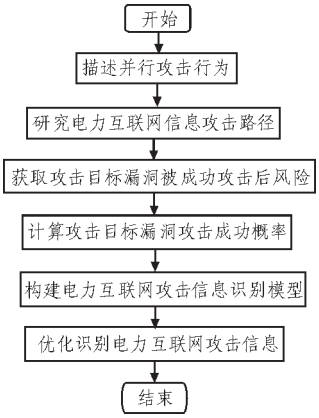

图3 攻击信息优化识别流程

4 实验验证

为了验证基于机器学习的电力互联网攻击信息识别方法，在网络整体安全性分析中的应用有效性，进行实验验证。

4.1 实验环境

实验环境中包括5台主机，20个网络节点，攻击者通过获取用户权限，并且攻击成功率可达到50%。实验环境如图4所示。

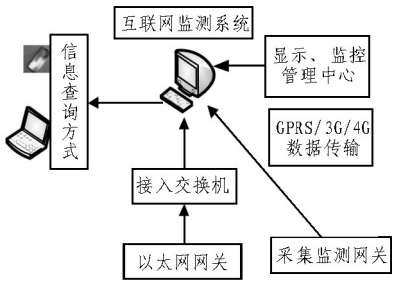

图4 实验环境

4.2 实验参数设置

实验参数设置如表1所示。

表1 实验参数设置

参量		数值
信息采集频率		40 kHz
网络信号窗	分窗大小	30维
	分窗长度	35 ms
	分帧	250点
	帧移	55点

为了防止电力互联网识别过程受到电力设备性能影响，需在图4所示实验环境中进行实验验证分析。

4.3 实验结果与分析

由于电力互联网中存在脆弱性潜在危险，分别使用基于综合权重法的电力互联网攻击信息识别方法、基于TOPSIS算法的攻击信息识别方法和基于机器学习识别方法对攻击信息进行识别，对比3种不同方法识别精准度。3种识别方法短时能量对比分析如图5所示。

在不同时间内，基于综合权重法和基于TOPSIS算法的识别方法会出现短时能量失效现象，而采用

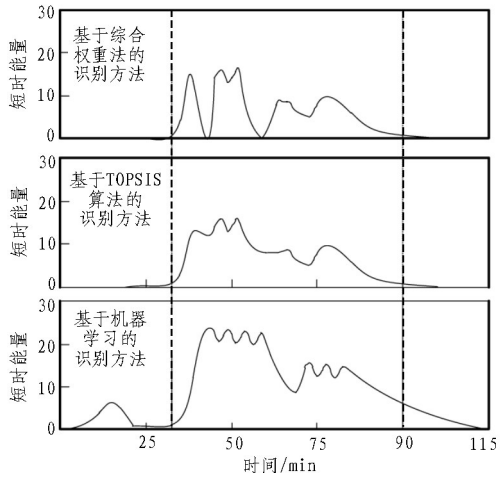

图5　3种方法短时能量对比分析

基于机器学习的识别方法不会出现短时能量失效现象,由此该方法可精准识别电力互联网攻击信息。

在短时能量下,对比3种方法分析并行攻击脆弱性,如图6所示。

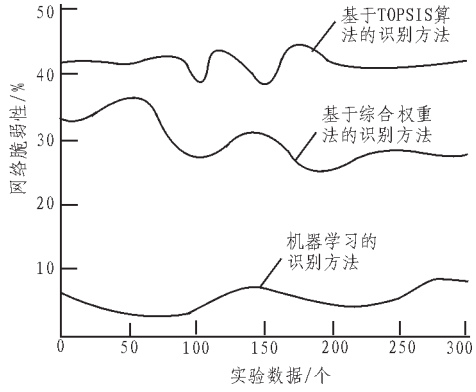

图6　3种方法并行攻击脆弱性

由图6可知,3种方法的网络脆弱性不同,基于机器学习的识别方法网络脆弱性保持在10%以下;基于综合权重法的识别方法网络脆弱性保持在30%~40%之间;基于TOPSIS算法的识别方法网络脆弱性保持在40%以上。因为基于机器学习的识别方法,通过确定电力互联网中任意安全漏洞被攻击选取为攻击接入点可能性指标,来对攻击信息进行识别,使得网络脆弱性较低。

为了进一步验证所提方法的合理性,需将3种方法识别精准度进行对比分析,结果如表2所示。

根据上述实验对比结果可知,3种识别方法中,基于机器学习的电力互联网攻击信息识别方法的识别精准度最高。

表2　3种方法识别精准度

实验次数/次	1	2	3	4	5
基于综合权重的识别方法	52%	67%	63%	62%	58%
	52%	61%	35%	50%	52%
	50%	54%	52%	50%	49%
	25%	28%	26%	31%	29%
基于TOPSIS算法的识别方法	65%	55%	69%	55%	45%
	54%	52%	65%	50%	42%
	52%	48%	42%	42%	38%
	38%	31%	30%	29%	20%
基于机器学习的识别方法	98%	98%	97%	97%	97%
	97%	96%	97%	95%	96%
	93%	92%	92%	93%	91%
	91%	92%	90%	90%	90%

5　结束语

为实现电力互联网攻击信息的高效识别,提出基于机器学习的电力互联网攻击信息识别方法。通过对电力互联网攻击信息识别和机器学习研究,能够适应互联网大数据的特点,并在此基础上选择机器学习识别方法,保证攻击信息识别误差达到最小。实验结果表明,该方法能够高效识别攻击信息。

参考文献:

[1] 董哲, 唐湘滟, 程杰仁,等.基于HMM时间序列预测和混沌模型的DDoS攻击检测方法[J]. 计算机工程与科学, 2018, 40(12):72-80.

[2] 马乐乐, 束永安. SDN环境下基于机器学习算法的DDoS攻击检测模型[J]. 微电子学与计算机, 2017, 35(5):21-26.

[3] 闻佳妍, 史昕昱, 邵杰,等. 基于综合权重法的电力系统脆弱线路识别方法[J]. 电工电气, 2019(6):10-14.

[4] 耿子惠, 崔力民, 舒勤,等. 基于TOPSIS算法的电力通信网关键节点识别[J]. 电力系统保护与控制, 2018, 46(1):78-86.

[5] 王世峰, 都凯悦, 孟颖. 基于机器学习的车辆路面类型识别技术研究[J]. 兵工学报, 2017, 22(8):189-195.

[6] 彭嘉宁, 常鹏, 张舒丽. 基于模拟技术的智能变电站远动信息快速校验方法应用[J]. 电子设计工程, 2018, 26(21):52-56.

[7] 赵海文, 齐恒佳, 王旭之. 基于机器学习的人机

(下转第74页)

离的计算次数，提升检索速度。同时，设计了两个可调参数来折衷精度与效率。最后通过在实际业务数据上的实验，验证了算法的实用性。

但算法也存在两点问题值得进一步探究：1）如何确定最优参考点个数和搜索宽度；2）是否有更优的参考点选择策略而非随机选取，来针对不同数据集选择具体的参考点。

参考文献：

[1] Chopra S，Hadsell R，Lecun Y. Learning a similarity metric discriminatively，with application to face verification[C]. IEEE Computer Society，2005.

[2] Parikh A，Das D. A Decomposable Attention Model for Natural Language Inference[C]. EMNLP. 2016: 2249-2255.

[3] Chen Q，Zhu X，Ling Z H，et al. Enhanced LSTM for Natural Language Inference[C]. ACL，2017: 1657-1668.

[4] Vaswani A，Parmar N. Attention is all you need[C]. NIPS，2017: 5998-6008.

[5] Devlin J，Lee K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. Association for Computational Linguistics，2019: 4171-4186.

[6] 谷重阳，徐浩煜，周晗. 基于词汇语义信息的文本相似度计算[J]. 计算机应用研究，2018，35(2): 530-536.

[7] Grigori S，Alexander G. Soft similarity and soft cosine measure[J]. Computación y Sistemas，2014，18(3): 125-136.

[8] Kusner M，Sun Y. From word embeddings to document distances[C]. ICML，2015: 957-966.

[9] Pele O. Fast and robust earth mover´s distances[C]. ICCV，2009: 460-467.

[10]Fu C，Cai D. Efanna:an extremely fast approximate nearest neighbor search algorithm based on knn graph[J]. arXiv，2016，(9): 7228-7248.

[11]唐明，朱磊，邹显春. 基于Word2Vec的一种文档向量表示[J]. 计算机技术，2016，43(6)：214-217.

[12]Mikolov T，Sutskever I. Distributed representations of words and phrases and their compositionality[C]. NIPS. 2013: 3111-3119.

[13]Mikolov T，Chen K. Efficient estimation of word representations in vector space[J]. arXiv，2013(1): 3781-3793.

[14]Bentley J L. Multidimensional binary search trees used for associative searching[J]. Communications of the ACM，1975，18(9): 509-517.

[15]Chen Z，Yan J. Fast KNN search for big data with set compression tree and best bin first[C].CCIOT. IEEE，2016: 97-100.

[16]Lv Q，Josephson W. Multi-probe LSH: efficient indexing for high-dimensional similarity search[C]. VLDB Endowment，2007: 950-961.

（上接第69页）

协调操作意图感知与控制方法研究[J]. 机床与液压，2019，22(10):66-68.

[8] 涂同珩，金炜东. 基于自动机器学习流程优化的雷达辐射源信号识别[J]. 计算机应用研究，2019，36(1):197-199.

[9] 林晓林，孙俊. 基于机器学习的小目标检测与追踪的算法研究[J]. 计算机应用研究，2018，35(11):256-259.

[10]胡峰松，李苍，王冕. 基于机器学习的SQL注入检测方案[J]. 计算机工程与设计，2019，22(6)：45-46.

[11]赵双，陈曙晖. 基于机器学习的流量识别技术综述与展望[J]. 计算机工程与科学，2018，40(10)：34-44.

[12]谭咏梅，刘姝雯，吕学强. 基于CNN与双向LSTM的中文文本蕴含识别方法[J]. 中文信息学报，2018，32(7):16-24.

[13]秦斐，梁兴东，张福博. 基于机器学习的阵列层析SAR建筑物目标提取方法[J]. 信号处理，2019，35(2):22-32.

[14]马媛媛，何高峰，张波. 电力资源网络对攻击信息优化识别仿真研究[J]. 计算机仿真2017，22(6):18-19.

[15]马唯唯. 移动互联网下信息采集安全加密研究[J]. 电子设计工程，2018，26(23):52-56.

[16]TANG Y. Anomaly Inference Based on Heterogeneous Data Sources in an Electrical Distribution System[D]. Houghton，MI，USA: Michigan Technological University，2018.