doi: 10.7690/bgzdh.2025.05.009

基于电网协同业务场景的数据全链路检测研究

谢　辉[1]，司福利[1]，张建中[1]，郑景立[1]，张继英[1]，陈飞云[1]，张　沛[2]
（1. 南方电网数字电网研究院有限公司，广州 510000；2. 天津弘源慧能科技有限公司，天津 300308）

摘要：为对电网协同业务场景中的数据全链路进行入侵检测，基于深度学习设计了全连接神经网络-决策树算法。对电网协同业务场景中存在问题产生的原因进行分析，对常见的分类算法进行介绍；将深度学习中的全连接神经网络和机器学习的决策树算法进行结合，得到所设计的全连接神经网络-决策树入侵检测算法模型。通过在相关数据上测试，该模型在网络入侵中的检测精准率可达 0.99，精度可达 0.984，召回率为 0.97，F_1 分数为 0.977。结果表明：该算法与同类算法相比优势突出，为电网协同业务场景中的数据全链路提供了更精准的技术支撑，对于电网规划具有重要的优化作用。

关键词：电网协同业务；全链路；入侵检测；电网规划；决策树；全连接神经网络
中图分类号：TM711　　**文献标志码：**A

Research on Data Full Link Detection Based on Power Grid Collaboration Business Scenario

Xie Hui[1], Si Fuli[1], Zhang Jianzhong[1], Zheng Jingli[1], Zhang Jiying[1], Chen Feiyun[1], Zhang Pei[2]
(1. *China Southern Power Grid Digital Grid Research Institute Co*., *Ltd*., *Guangzhou* 510000, *China*;
2. *Tianjin Hong Yuan Smart Energy Company Limited*, *Tianjin* 300308, *China*)

Abstract: Based on deep learning, a fully connected neural network-decision tree algorithm is designed for intrusion detection of full data link in power grid collaborative business scenarios. This paper analyzes the causes of the problems in the power grid collaborative business scenario, and introduces the common classification algorithms. It combines the fully connected neural network in deep learning with the decision tree algorithm in machine learning, and obtains the designed fully connected neural network-decision tree intrusion detection algorithm model. Through the test on the relevant data, the detection accuracy of the model in network intrusion can reach 0.99, the precision can reach 0.984, the recall rate is 0.97, and the F_1 score is 0.977. The results show that compared with similar algorithms, the proposed algorithm has outstanding advantages, provides more accurate technical support for the full data link in the grid collaborative business scenario, and plays an important role in the optimization of power grid planning.

Keywords: power grid collaborative business; full link; intrusion detection; power grid planning; decision tree; fully connected neural network

0　引言

在互联网高度发展与深度应用的背景下，已真正进入大数据时代。数据会利用数据链进行交互，但与此同时，网络安全在很多应用场景中受到了越来越多的关注。电力物联网是围绕电力系统各环节，充分应用移动互联、人工智能等现代信息技术，实现电力系统各环节万物互联、人机交互，具有状态全面感知、信息高效处理、应用便捷灵活特征的智慧服务系统。区域协同电网规划应以“横向协同，纵向联动”为主旨思想，建立科学、安全、高效、优质的电网规划。由于电网规划中信息数据较为复杂，现有智能决策平台实时性较差，对电力系统决策水平造成影响。电网规划发展带来更密集的网架和更复杂的系统，在开展设计工作时，需要收集更详细的基础数据资料，各专业之间的协同设计会更加密集[1]。为有效解决电网规划的问题，基于协同业务的电网规划应运而生。

“全链路数据”是对多源、多维信息的获取、表示及其内在联系进行综合处理和优化，并形成完整准确、及时和有效的综合信息。数据在数据链中传输、计算所产生的“全链路”，往往也伴随着数据安全的问题，近几年深度学习凭借其强大的处理能力，为数据的“全链路”入侵检测提供新的角度。神经网络是模仿人脑进行信息处理的算法，具有强大的自学习、自适应、非线性匹配和信息处理能力。将神经网络技术应用在数据融合中，可有效减少冗

收稿日期：2024-08-15；**修回日期：**2024-09-12
第一作者：谢　辉(1986—)，男，广东人，硕士。

余数据传输，提高“全链路”数据融合的实时性与精度，改善数据融合算法的性能。

电网门户网站承载的信息化系统较多，比如电力物资综合采购、网上营业厅、电力营销双向互动平台、电力交易市场、企业邮箱等[2]。这些系统通过互联网为普通用户提供电力信息发布、自主缴费等便民服务。作为国民生活的基础设施服务方，其外部网站的安全稳定运行具有重要的社会影响。由于电网系统潜藏了许多黑客、木马和病毒等攻击威胁，严重阻碍电网互联互通，因此会对电网协同业务造成影响。为提高电网协同业务中对电网网络的数据分析能力，可以采用先进的大数据分析技术，实现电力网络状态检测、入侵检测，并且构建一个监测报警系统，将检测到的威胁发送给防御系统，从而实现电网门户网、局域网和互联网之间的通信安全。在人工智能广泛使用之后，将卷积神经网络用到了入侵检测问题中，实际的应用效果也较为理想。但是在大数据的时代，网络运行中的数据量传输速度和传输数量非常庞大。对于传统的深度学习或机器学习方法而言，其在入侵检测问题中的准确率还需要进一步优化，从而适应时代的发展[3]。

目前，智能电力系统主要由一系列承担不同角色的智能嵌入式电力终端组成，如配电终端单元、变压器终端单元、馈线终端单元等。这些智能嵌入式终端通过与智能电网的互动传输电力信息，使整个网络更加智能化。然而目前智能电力系统主要存在 2 个问题：1) 电力信息复杂且数据量大，常规的入侵检测系统难以应对如此繁重的计算压力；2) 在引入大量异构电力智能终端设备的同时，这些设备本身也存在大量漏洞，容易被攻击者利用，成为进一步攻击电网主站的跳板，一旦电力智能终端被入侵，整个电网将面临被破坏的可能。针对现有研究的不足，笔者设计了全连接神经网络-决策树算法作为电网系统的入侵检测方法。相关数据测试结果证明，该模型在网络入侵中的检测精准率和精度都有较好的优势，可为电网协同业务场景中的数据全链路提供更加优秀的检测算法支撑，对于电网规划来说具有重要的优化作用。

1 研究方法

1.1 电网协同场景中的问题

电网规划是电力系统建设中非常关键的内容，其规划的科学与否，最终会对电网投资收益和电网安全产生重要影响。而配电网通常分布十分广泛，其所牵涉的设备也较多，这为相关的管理带来很多不变。目前电网规划的挑战如图 1 所示。

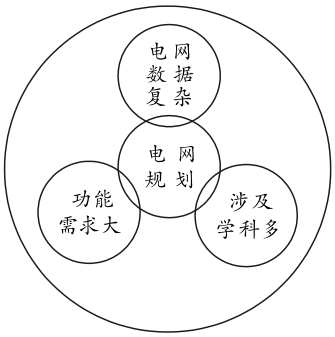

图 1 电网规划的挑战

在电网规划的各种挑战中，如图 1 所示有 3 个核心要点；1) 由于电网数据非常的繁琐，数据量也非常的庞大，这就需要比较统一的管理平台对其进行管理；2) 电网会对很多专业学科均有所设计，这就需要很多不一样的专业技术；3) 对电网问题的分析常常会从很多角度出发，功能的复杂性，导致很难获取到电网的第一手资料，从而对电网的重要指标不能掌控，很难实现规范化[4]。对电网进行规划很难靠人工单方面进行分析处理，电网规划与电网运行数据以及信息化平台密不可分[5]。为了电网的发展，往往会基于业务协同进行电网规划，从而能够拥有智能化决策的能力。改变了以往的电网决策方式，使电网规划和决策速度得到了有效升级，在电网协同业务场景中，数据成为非常重要的资源，数据在网络中利用数据链路交互，网络安全也因此产生[6]。

电网网络为企业办公带来便利的同时也带来了黑客、病毒和木马等攻击威胁，这些威胁可以盗取广西电网客户信息、用户缴费账号信息等资源，破坏电网网络的正常运行。因此，采用先进的大数据挖掘分析技术构建完善的网络安全监测报警系统，可以集成多种网络安全防御技术，包括入侵监测、状态检测等。实现一个融合、主动的网络安全防御系统，可进一步提升电网网络的安全防御能力。为电网协同业务场景中的数据全链路进行检测，把深度学习的全连接神经网络和以往机器学习的办法进行结合，对数据全链路进行入侵检测，让电网数据链路中的安全得到保障。

1.2 电网协同业务中的数据处理分类算法

在电网协同调度的业务中，主要借助状态估计和参数辨识来提升电网数据参数。业务系统对数据分析处理的实时性要求及业务间资源共享协同性的

要求越来越高；通过业务间数据共享协同来实现对电力系统海量的历史数据进行深度分析挖掘的需求也越来越多。在数据处理分类算法的作用下，能对各个电网网点的运行状态进行有效地监控，以便根据这些电网数据对电网实施优化且高效地协同与调度。

智能电网中负荷预测、设备异常检测、视频监控等场景下，实时性要求高，数据量大，容易产生网络拥塞，需要数据挖掘、数据压缩等技术的支持。所谓的分类就是以特定的训练样本数据作为依据，构建相关的分类函数，通过所组建的函数模型去对需要未经过处理的数据进行相关分类，从而能够快速达到目的[7-8]。以下是使用比较频繁的分类算法。

1) 决策树算法。

决策树为树状结构，它从根节点开始，对数据样本进行测试，根据不同的结果将数据样本划分成不同的数据样本子集，每个数据样本子集构成一子节点。决策树通过一系列规则对数据进行分类，提供一种在特定条件下会得到特定值的类似规则的方法。决策树分为分类树和回归树 2 种，分类树对离散变量做决策树，回归树对连续变量做决策树。一般的数据挖掘工具，允许选择分裂条件和修剪规则，以及控制参数来限制决策树。决策树作为一棵树，树的根节点是整个数据集合空间，每个分节点是对一个单一变量的测试，该测试将数据集合空间分割成 2 个或更多块。每个叶节点是属于单一类别的记录。图 2 表示的是重要的决策树算法。

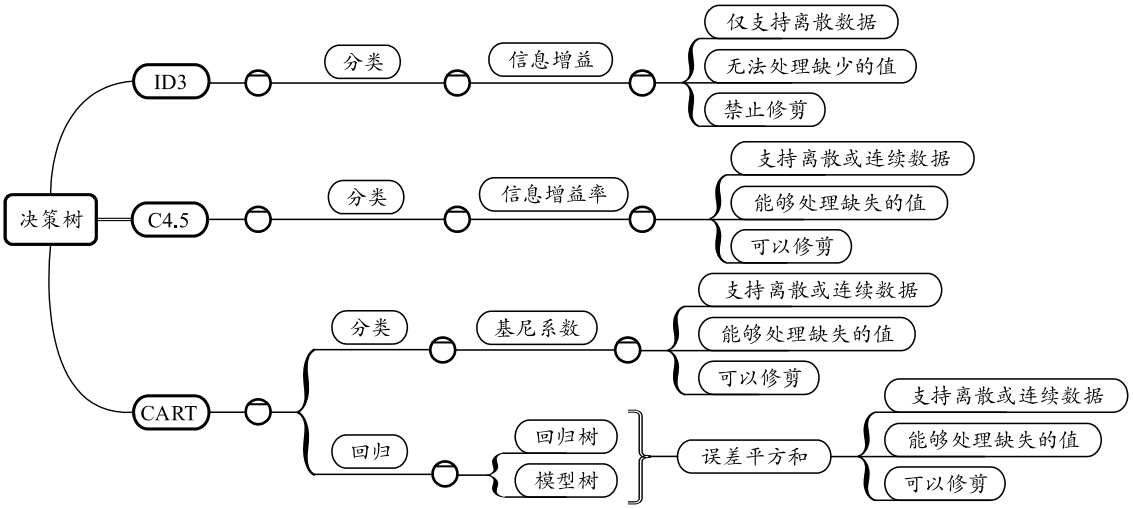

图 2　重要的决策树算法

2) 基于贝叶斯的分类算法。

分类的目的是提出一个分类函数或分类模型。该模型能把数据库中的数据项映射到给定类别中的某一类。通过分析训练数据样本，产生关于类别的精确描述，可用来对未知的数据进行分类预测。分类算法的核心部分是构造分类器。在众多分类方法和理论中，朴素贝叶斯由于计算高效、精确度高，并具有坚实的理论基础而得到了广泛应用。它运用的分类原理是运用贝叶斯公式按照某个对象的先验可能性推导出其后验的可能性，让后验可能性最大的类充当这个对象所属的类[9]。

1.3　全连接神经网络算法的设计

深度神经网络是深度学习中的经典网络，其具备超强的数据表征性，可对数据特征进行自动提取；因此，该网络十分适用于电网系统中的数据特征提取和挖掘，能够实现态势精准感知和预测。全连接神经网络和决策树算法的设计，需要对检测数据进行预处理，把数据利用特征映射和数字归一化得到所需要的数据集。进行数据预处理的目的在于将类别各异以及特征异常的数据转换为标准统一的数据[10]。处理后的数据，能够使特征提取和学习训练更加高效，从而奠定了数据基础，使最终的检测效果更加优秀。

要实现抽象特征提取，需要设定一个特定的模块；因此，必须基于 3 个不可或缺的参数，实现全连接层的构建，并以此为基础来实现模块功能，服务于特征提取[11]。不同于其他全连接神经网络，这类模型所构建的全连接层，是以激活函数为内核进行抽象特征的获取，没有用 softmax 分类器进行分类笔者所提模型未直接利用浅层分类模型进行分类，是由于从抽象特征和未转换的数据特征对比来看，抽象特征可以更加优秀地展现。其可以剔除掉

不存在或极低现实价值的信息，把运用价值比较大的提取出来，从而使检测的效果得到优化。不采用 softmax 的原因是因为浅层分类模型的检测能力在入侵检测中性能要更加良好。综合多种因素，笔者使用了特殊网络，同时引入了分类器。图 3 展示了基于全连接神经网络基础的抽象特征提取流程。

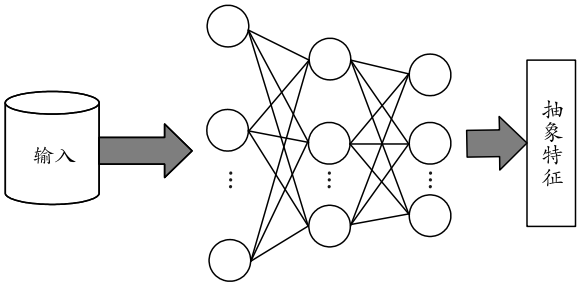

图 3 基于全连接神经网络基础的抽象特征提取流程

在全连接神经网络基础的抽象特征提取流程中，重点把握提取抽象特征的环节。聚焦全连接神经网络，在最左侧区域，承载的任务是数据输入；而在中间区域，这些节点承载的是分析与求解；最右侧区域，承载的任务是输出数据。对于中间节点的求解，将 sigmoid 作为激活函数。具体到求解方面，单一节点的运算需先执行线性求解，之后通过激活函数实现非线性求解。具体如下：

$$F(x_i)=a(z_i)=\text{sigmoid}(w_i*x_i+b)；\tag{1}$$

$$a(z_i)=\text{sigmoid}(z_i)；\tag{2}$$

$$z_i=w_i*x_i+b。\tag{3}$$

式中：z_i 为线性分析；$F()$为节点整体运算；$a()$为激活函数为支撑的非线性分析；w_i 为权值；b 为参数；x_i 为最后的输出。在网络层条件下，每个节点都应按照如上规范来求解，上一层的输出结果，具体化为下一层的输入结果，直至得到最终参数。通过该方法来提取抽象特征，可以为之后的训练分裂奠定基础。

全连接神经网络将输入数据作改变，得到抽象数据，然后借助 softmax 分类器，得到浅层分类模型，并据此对入侵检测进行分类。笔者选用决策树来替代浅层分类模型。决策树的结构与树形状的流程图比较相近，决策树各个节点是按照特征分析所得数据，实现步的分类处理。分支节点是基于信息熵给出的，信息熵一般是来分析数据量化相关内容的。把握特征，对其存在的信息增益进行分析，找出信息增益显著特征，并将该部分设定为分裂节点。

搭建决策树前首先要针对划分的抽象特征数据集进行分析，求解出具体的信息熵；其次，针对尚未划分的抽象特征，完成其关联信息熵的界定；找出信心增益最突出的特征，并据此对数据进行划分；最后，递归处理划分之后的子抽象特征数据集，找出最优数据，以此为条件完成分子抽象特征数据集的划分。取得抽象参数，将其关联到决策模型，在进行相关的训练之后，进行最后的分类。

综上，笔者采用基于 C5.0 算法的决策树模型，经过数据采集、属性指标值、训练样本抽取，建立评估初始决策树。然后，通过调整决策树的修剪程度、确定最佳成本矩阵、寻求最佳 Boosting 迭代次数确立最优决策树的相关参数，形成最优决策树。将最优决策树提取出来的重要性排序较高的指标输入神经网络，最终输出分类结果。

1.4 案例分析

为测试笔者所提算法的性能和可行性，试验平台为：Windows 10 旗舰版 64 位，处理器 11th Gen Inter(R)Core(TM)i5−1135G7@2.40 GHz，内存 16 G。软件环境用到 Jdk1.8、Eclipse2019、Python3.7。

采用的实验数据集为：数据挖掘和知识发现（data mining and knowledge discovery，KDD）Cup 99 数据集。该数据集是 1998 年由麻省理工学院林肯实验室为入侵检测模型评估而建立的测试数据集，该数据集一共 490 万条记录，每条记录共有 41 个数据属性和一个标识属性。41 个数据属性描述数据的特点，除了一些基本属性外，还利用领域知识扩展了一些属性，一个标识属性描述数据是否为入侵行为。KDD Cup 99 数据集中共包含 24 种类型的攻击，根据攻击方法和目的可以分为 4 类：PROBE、DOS、U2R 和 R2L。PROBE 为进行信息收集的攻击类型；DOS 为拒绝合法用户请求的攻击；U2R 为远程主机非法获取本地主机权限的攻击；R2L 为本地非超级用户获取超级用户权限的攻击。

在模型训练时间，将 KDD Cup 99 数据集中正常数据按照 7:3 划分，70%的正常数据作为训练集数据，另外 30%作为测试集的正常数据。测试集为检测不同类型的异常，设置正常数据和异常数据的比例在 1:1 左右。然后，在独热编码阶段，采样数据的维度会增加，因为字符类型的特征被替换成数字特征。对于所收集的数据集，字符类型的特征是协议类型、标志和服务。在独热编码之后，特征从 41 维变成 118 维；最后，对所有的数字特征进行归一化，使其取值全部处于同一范围内。

2　研究结果

2.1　不同入侵检测算法模型的准确度和精度测试

将数据集里面的数据选取 0.1%的数据作为本次实验的训练集、验证集、测试集，不同入侵检测算法模型的准确度和精确度结果如表 1 所示。

表 1　不同入侵检测算法模型的准确度和精确度

算法	准确度	精确度
决策树	0.92	0.94
朴素贝叶斯	0.90	0.93
随机森林	0.91	0.92
全连接神经网络-决策树	0.99	0.98

经过对不同算法模型的入侵检测准确度和精确度对比，优化的全连接神经网络-决策树算法模型的检测准确度可达到 0.99，精确度可达到 0.98。相较于其他 3 种算法而言，从检测准确度和精确度角度来看，全连接神经网络-决策树算法是最优的。

2.2　各检测算法召回率和 F_1 分数对比

本次实验，依旧是将数据集里面的数据选取 0.1%的数据作为本次实验的训练集、验证集、测试集，得到不同入侵检测算法模型的召回率和 F_1 分数对比情况，结果如表 2 所示。

表 2　不同检测算法召回率和 F_1 分数

算法	召回率	F_1
决策树	0.94	0.93
朴素贝叶斯	0.91	0.94
随机森林	0.92	0.93
全连接神经网络-决策树	0.97	0.98

从表 2 的召回率和 F_1 分数可以看到，笔者所提全连接神经网络-决策树模型的召回率达到了 0.97，F_1 分数达到了 0.98，相较于比较的其他算法，数值是最高的。综合 2 个实验来看，笔者所提全连接神经网络-决策树入侵检测算法的检测性能较优秀，在基于协同业务对电网规划时，可以对它数据全链路里面的安全隐患进行很高的检测，提升电网规划的安全性。

2.3　负荷损失率与线路容量极限的关系

利用度-度耦合系统，在随机攻击策略下，对比不同线路容量极限下的平均负荷损失率。将电力线路容量分别设置为初始线路的 2、3、4 倍进行分析。结果如图 4 所示。

图 4 中，L=2, L=3, L=4 分别表示电力线路容量为初始线路的 2、3、4 倍。在信息节点损失数量达到 30 个时，系统完全崩溃。主要是随着线路容量极限的增加，当电力通信节点遭到攻击时，不能全局调节负荷，电力元件超过极限时会导致系统故障蔓延，增加线路的极限，会降低超过极限的概率，减少负荷损失率，增强鲁棒性。对比不同容量极限的攻击节点数量发现，在线路超过 3 倍的初始容量之后，容量极限增加，每个元件的最大承受负载能力相应提高。

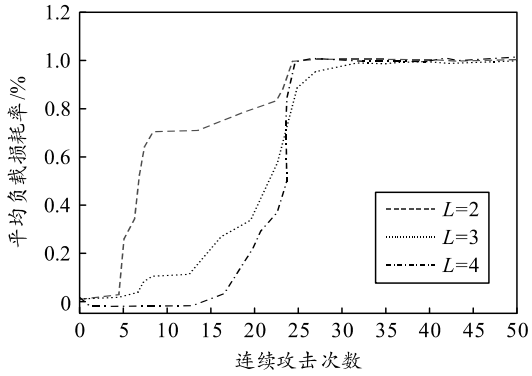

图 4　不同线路容量极限下的平均负荷损失率

3　结论

笔者基于深度学习，设计了全连接神经网络-决策树算法，从而实现对数据全链路中的入侵进行检测。通过在相关数据上测试，设计的全连接神经网络算法模型在网络入侵中的检测精准度和精确度都有很好的比较优势，在入侵检测上可以拥有很好的实际效果。笔者所提数据全链路检测系统遵循模块化原则、可靠性原则以及扩展性原则。通过对涉及生产专业的关键业务流程进行需求分析，结合全连接神经网络-决策树入侵检测模型，重点分析探讨了模型与数据双重驱动背景下智能电网信息物理协同安全防护的理论方法，为高度智能化、自动化的智能电网安全防护奠定理论基础。

参考文献：

[1] 石城, 雷海. 智能电网对于现代电力系统的需求与响应研究[J]. 电子乐园, 2022(12): 130-132.

[2] 刘新, 常英贤, 孙莉莉, 等. 基于深度学习的电网网络攻击检测研究[J]. 自动化仪表, 2022, 43(12): 81-85.

[3] 杜露露, 石倩倩, 王有军. 基于人工神经网络的新能源电网微变检测方法[J]. 电器工业, 2022(11): 55-58.

[4] 杨凤霞. 云计算环境下的智能电网光通信网络路由算法[J]. 自动化与仪器仪表, 2017(9): 24-27.

[5] 李学龄, 崔焱, 柴雁欣. 面向电网协同业务场景的数据全链路检测研究[J]. 自动化仪表, 2022, 43(12): 49-52, 57.

（下转第 51 页）