计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241200157

基于多源网络数据的电力监控系统入侵检测方法

蒋亚坤　林　旭
云南电网有限责任公司云南电力调度控制中心　昆明 650000

摘　要　随着电力系统信息化、网络化、智能化建设的不断推进,电力监控系统面临着日益严峻的网络安全威胁。综合考虑网络资产安全风险、用户行为、业务特征等多方面因素,对电力监控系统网络所涵盖的多源数据进行全面深入分析显得尤为重要。据此提出了电力监控系统多源数据清洗方法及入侵检测方法,利用改进最大相关-最小冗余算法对电力监控系统网络的多源安全数据特征进行选择,保留合适的安全数据特征,利用网络入侵检测模型实现多源网络安全数据的检测与分类,有效解决电力监控系统多源数据特征属性复杂导致后期模型分类准确度下降等问题。仿真实验证明,所提出的多源数据特征选择方法与入侵检测算法对电力监控系统攻击的检测率与分类准确率均有明显提高。

关键词:电力监控系统;长短期记忆神经网络;特征处理;入侵检测;数据清洗
中图分类号　TP309

Intrusion Detection Method for Power Monitoring System Based on Multi-source Network Data

JIANG Yakun and LIN Xu
Yunnan Power Grid Corporation Yunnan Power Dispatch Control Centre,Kunming 650000,China

Abstract　With the continuous advancement of informatization,networking,and intelligence in the power system,the power monitoring system is facing increasingly severe network security threats. It is particularly important to conduct a comprehensive and in-depth analysis of the multi-source data covered by the power monitoring system network,taking into account various factors such as network asset security risks,user behavior,and business characteristics. Based on this,a multi-source data cleaning method and intrusion detection method for power monitoring systems are proposed. The improved maximum correlation and minimum redundancy algorithm is used to select the multi-source security data features of the power monitoring system network,retain appropriate security data features,and use a network intrusion detection model to detect and classify multi-source network security data,effectively solving the problem of complex feature attributes of multi-source data in power monitoring systems leading to decreased accuracy of model classification in the later stage. Simulation experiments show that the proposed multi-source data feature selection method and intrusion detection algorithm have significantly improved the detection rate and classification accuracy of attacks on power monitoring systems.

Keywords　Power monitoring system,Long short-term memroy,Feature processing,Intrusion detection,Data cleaning

1　引言

电力监控系统网络的多源安全特征是保障整个系统安全和稳定运行的重要组成部分,主要由网络资产安全风险、用户行为、业务特征等多方面因素构成,涉及到电力监控系统的网络设备、传感器、通信链路等各种资产的安全状况,网络设备的漏洞情况、补丁更新状态、访问控制策略、关键设备和系统的实时监控和异常检测等方面。

Song 等基于原始测量数据集,利用双向长短期记忆(BILSTM)预测的改进插值方法进行电力系统的多源数据处理,实验证明与传统方法相比,该方法能够更准确地区分不同类别数据,同时提高计算的精度和速度[1]。Wu 等提出了一种基于多源异构数据融合的电力系统安全控制方法,通过实验验证了该方法的有效性[2]。Sahu 等利用网络测试平台执行多源数据融合以训练入侵检测数据集,并将数据合成为算法的特征以检测入侵,实验证明,融合来自多个数据源的信息可以帮助识别网络诱发的事件并减少误报[3]。Guo 等提出了一种基于多源异构数据关联规则分析的识别方法,利用 Bi-GRU 网络进行相关性和冗余分析以提取指标的权重,最后,通过算例验证了该方法的有效性[4]。He 等为了满足海量数据实时处理需求,实现多源、多位置、多时态数据的快速转换,提出了一种针对海量多源数据的电力监控系统架构,用于解决海量数据快速提取、分析、存储问题,实现终端行为状态监控和网段流量态势实时感知[5]。Jiang 等为了解决电力系统中终端设备易产生大量异构数据而给系统带来巨大的数据负载压力等问题,提出了一种基于欧氏距离加权优化的异构数据边缘侧处理算法,计算了多源异构数据之间的相似性,以消除数据冗余[6]。Park 等基于多源的真实操作数据提出了一种特征提取方法,通过实验验证了该方法的有效性[7]。Li 等对网络攻击行为进行了研究,建立了具有迁移学习能力的攻击行为的数学统计分析模型,开展了网络入侵大数据融合分析、多源异构数据融合分析等技术的研究[8]。Dong 等基于多

基金项目:中国南方电网有限责任公司科技项目(0500002023030301XT00152)
This work was supported by the China Southern Power Grid Co.,Ltd.(0500002023030301XT00152).
通信作者:蒋亚坤(15804324722@163.com)

源数据和一维轻量级卷积神经网络提出了一种检测方法，提高了数据的利用率，并通过实验证明了所提方法在多源数据融合中的性能[9]。Xiong 等提出了联邦生成模型框架，解决了多源异构数据在特征相关场景和标签相关场景两种场景下的分布式数据生成问题[10]。Ankitdeshpandey 等研究了机器学习算法在检测和识别电网攻击方面的适用性，构建深度神经网络(DNN)模型以衡量 DNN 在网络攻击检测方面的有效性。测试结果证实，SVM、随机森林和 DNN 算法适用于在电网系统上部署入侵检测系统设备[11]。Han 等为了解决电力系统测量数据具有高维特征和强噪声、难以直接用于入侵检测等问题，提出了一种新型的二元粒子群包裹特征选择优化框架(BPSWO)，该框架可以通过加强特征选择与训练之间的耦合来提高机器学习方法的入侵检测精度[12]。Dasgupta 等针对电力系统资产的网络攻击进行了总结，从攻击的特点、受影响设备等方面进行了阐述[13]。Aljohani 等提出了一种利用深度学习神经网络(DLNN)的入侵检测和缓解系统(IDMS)来检测、分类和定位电力监控系统中的入侵行为，仿真结果表明，所提检测、分类、定位和预测方法具有较高的精度[14]。

以上方法在电力监控系统多源数据处理方面虽取得了一定的进展，但在数据融合效率、多源异构数据特征选择、异常值处理方面还有待完善和提高[15-16]，据此通过综合考虑网络资产安全风险、用户行为、业务特征等多方面因素，可以有效提升电力监控系统的安全性。因此，提出了基于最大相关最小冗余的多源异构数据特征选择方法，以深入分析电力监控系统网络的多源安全数据特点，同时考虑特征与目标变量的相关性以及特征之间的冗余性，避免只选择相关性高且冗余性高的特征，进而设计了针对性的网络入侵检测模型以实现对各种类型电力监控系统网络入侵的全面检测与有效识别。

2 研究背景

针对多源网络安全特征包含的网络资产安全风险、用户行为、业务特征等安全要素，利用网络公开数据集分析多源网络安全特征[17]，提出电力监控系统多源异构数据清洗方法，用于实现对原始数据的清洗、整合、转换和优化，以提高数据的质量和适用性，为后续的数据分析和建模工作奠定基础。

2.1 多源网络安全数据集

考虑到电力监控系统爆发式海量数据出现的频率大幅增加，且这些数据具有很强的随机性，对实时性要求高。为了让所提方法具有一定的通用性和可验证性，利用网络公开且与电力监控系统具有强相关的数据集构建电力监控系统入侵检测数据集，包括密西西比大学网络攻击数据集和 UNSW-NB15 数据集[18]，同时在数据选择时重点考虑电力监控系统中可能出现的攻击数据类型。

2.1.1 密西西比大学网络攻击数据集

密西西比大学网络攻击数据集采用百分之一的比例进行随机采样而得，其中包含 4 种攻击类型：指令注入攻击、响应注入攻击、拒绝服务(Denial of Service，DoS)攻击、侦察攻击。数据来源的框架配置如图 1 所示，其中 G1 和 G2 表示发电机；R1，R2，R3，R4 表示智能电子设备，也称 IED，它的功能是可以控制断路器，断路器用 BR 表示，它们的编号为 1 到 4，框架中共包含两条线路，第一条是从 BR1 到 BR2，第二条是从 BR3 到 BR4。每个智能电子设备分别控制一个断路器，对应 IED 的编号就是控制断路器的编号。另外智能电子设备 IED 使用了距离保护机制，在检测到故障时，不管该故障是有效的故障还是伪造的故障，它都会控制断路器跳闸。此外，操作员可以手动跳闸断路器，一般是在对线路进行维护和升级时进行手动控制。

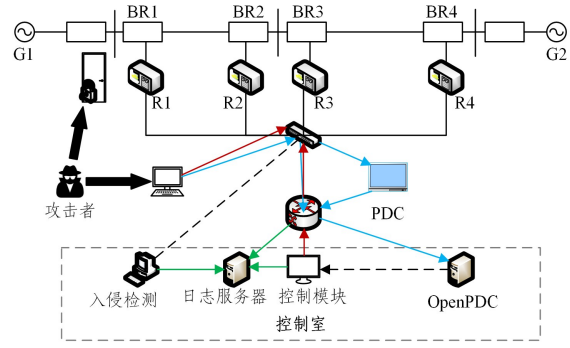

图 1 密西西比大学电力攻击数据集网络拓扑

Fig. 1 Network topology of the power attack dataset from the University of Mississippi

2.1.2 UNSW-NB15 数据集

UNSW-NB15 数据集包含 9 种攻击类型，分别为：

1)模糊攻击。攻击者首先向应用输入大量随机数据(模糊)让应用崩溃。然后用模糊测试工具发现应用的弱点。如果目标应用中存在漏洞，攻击者即可展开进一步漏洞利用。

2)分析攻击。包含不同的攻击端口扫描，垃圾邮件和 html 文件渗透。

3)后门攻击。一种秘密绕过系统安全机制来访问计算机或其数据的技术。

4)拒绝服务攻击。攻击者占用大量内存资源而使合法用户无法访问计算机或处理正常请求。

5)漏洞利用。攻击者知道操作系统或软件中的安全问题，并通过漏洞来利用这些知识。

6)通用攻击。适用于所有块密码(具有给定的块和密钥大小)，而不考虑块密码的结构。

7)侦察攻击。包含所有可以模拟收集信息的攻击。

8)外壳代码攻击(Shellcode)。在利用软件漏洞时用作有效载荷的一小段代码。

9)蠕虫攻击。攻击者使用计算机网络来传播、依靠目标计算机的安全故障访问、可以自我复制的攻击。

表 1 不同攻击类型在数据集中的数目

Table 1 Number of attack types in the datasets

类别标签	UNSW-NB15 训练集	UNSW-NB15 测试集
模糊攻击	995	9711
分析攻击	45927	7458
后门攻击	11656	2421
漏洞利用	52	200
通用攻击	2390	1009
侦察攻击	7865	230
外壳代码攻击	2852	149
蠕虫攻击	23	232
合计	125973	22544

建立训练集与测试集的步骤如下：

首先根据以上两种数据集，以 30%和 10%的概率选取攻击记录分别放入训练集与测试集。

其次，向测试集中放入另外未经训练的 3 种攻击类型，包括：

1）僵尸网络（Botnet）攻击：攻击者传播“僵尸程序”以形成“僵尸网络”，达到控制电脑、下发远程指令的目的。

2）分布式拒绝服务攻击（Distributed Denial of Service，DDOS）：攻击者以网站和服务器为目标，中断网络服务、耗尽应用程序资源，使合法用户无法访问计算机或处理正常请求。

3）网页（Web）攻击：攻击者通过漏洞扫描器扫描网站，对易受攻击的网站进行不同类型的网页攻击。

最后，放入与攻击记录总数相等的正常记录，使数据集中的正常与攻击记录数量分布均匀。

2.2 多源网络安全数据预处理

由于初始数据集分别来源于不同的网络公开数据集，在不同的场景下采集实现，需要形成一个攻击种类多样的电力监控系统入侵检测数据集，并且数据量整体较大，不同特征的取值规模不一样，为不影响模型训练，需要先进行标准化，所以利用上述网络公开数据集构建电力监控系统入侵检测数据集，比较特征属性间关系，对其进行包括数据集成、缺失数据检查、列名检查、数据转换等在内的数据预处理操作。

2.2.1 数据集成

对两种特征相同但攻击类型不同的数据集进行合并，形成具有攻击种类多样、覆盖范围广的电力监控系统多源网络入侵检测数据集，数据集的具体信息如表2、表3所列。

表2 入侵检测数据集-训练集

Table 2 Intrusion detection dataset-training set

类别标签	样本数量
模糊攻击	45927
分析攻击	995
后门攻击	45927
漏洞利用	11656
通用攻击	52
侦察攻击	2390
外壳代码攻击	7865
蠕虫攻击	2852
Benign	23
合计	436283

表3 入侵检测数据集-测试集

Table 3 Intrusion detection dataset-test set

类别标签	样本数量
Botnet	28169
DoS	7458
DDOS	70161
Web	5794
模糊攻击	9711
分析攻击	7458
后门攻击	2421
通用攻击	1009
侦察攻击	230
外壳代码攻击	149
蠕虫攻击	232
漏洞利用攻击	200
Benign	93462
合计	226454

2.2.2 缺失数据检查

数据集中的数据可能存在缺失值，会影响模型的准确性和稳定性。检查每条数据记录的每个维度是否存在空值并统计空值数量，若存在空值且空值数量远远小于总体数据样本量，则对存在空值的样本进行过滤，采用无NAN值样本。

2.2.3 字符型特征数值化

数据集中含有字符特征（特征“协议类型（protocol_type）”、特征“服务类型（service）”、特征“标志位（flag）”和特征“攻击类型（attack_type）”），但因深度学习中数据的数据大多是数字矩阵的方式，且字符型特征数值化可将复杂的字符信息简化为数值形式，降低模型的复杂性，增强模型的鲁棒性，使决策过程更加清晰，提高模型的可解释性，因此需要将非数值型数据转换为数值型数据。采用独热编码方式将字符特征转化为数值型特征。例如，特征“协议类型（protocol_type）”包括TCP，UDP和ICMP 3种状态。通过独热编码，将它们分别表示为001，010和100。类似地，另一个特征“服务类型（service）”包含70种状态，以及特征“标志位（flag）”有11种状态，利用独热编码的方式将它们转换为数值型数据。

2.2.4 数值特征归一化

将数据转换为特定范围或特定的分布，可以提高模型的收敛速度和稳定性。利用最大-最小归一化方法对各特征列进行处理，每个特征根据式（1）线性映射到[0，1]范围：

$$x_i=\frac{x_i-Min}{Max-Min} \tag{1}$$

其中，Max表示每个特征的最大值，Min表示每个特征的最小值，x_i表示第i个特征的数值。经过以上数据处理后，形成了电力监控系统网络攻击的多源异构数据集。

3 基于mRMR的多源异构数据特征选择方法

最大相关最小冗余（mRMR）是一种特征选择算法，它通过计算每个特征与目标变量之间的相关性，以及各个特征之间的冗余程度，确定哪些特征对于任务最为关键。它的目标是选择与目标变量高度相关、彼此之间具有最小冗余的特征集合。通过这种方式，mRMR算法可以帮助提高特征选择的效率和模型的性能，同时简化模型的复杂度，从而在保持高效的同时取得较好的分类结果。

为了优化算法的分类效能并精简特征集，针对入侵检测数据集中普遍存在的高维数据冗余及样本特征间高度关联的问题，提出了基于mRMR的多源异构数据特征选择方法。通过构建特征选择模型，巧妙地结合了相关系数法与多距离加权函数，对网络流量数据包中的特征进行细致筛选，有效剔除了冗余且高度相似的特征，从而保持了对目标变量的高度表征能力。

首先，采用皮尔逊（Pearson）相关系数法对攻击类型与攻击类型特征进行相关性分析，Pearson相关系数是用来描述两个变量之间的线性关系的关系变量，可以量化两个变量之间的相关性程度，为确定关键特征提供具体的数值依据。通过设定一定的阈值，可以筛选出与攻击类型具有较强相关性的特征属性，从而提高网络安全攻击检查的准确性和效率。计算式如下：

$$\rho_{x,y}=\frac{\left|N\sum_{i,j=1}^{N}x,y-\sum_{i=1}^{N}x\sum_{j=1}^{N}y\right|}{\sqrt{N\sum_{i=1}^{N}x^2-(\sum_{i=1}^{N}x)^2}\sqrt{N\sum_{j=1}^{N}y^2-(\sum_{j=1}^{N}y)^2}} \tag{2}$$

其中，相关系数$\rho_{x,y}$表示第i个攻击类型特征x_i与第j个攻击类型特征y_j之间的关系，相关系数的范围在-1到1之间，越接近-1或1表示两个变量之间的线性关系越强，越接近0则表示相关性越弱；N为入侵检测数据集中数据记录的个数。

其次，利用多距离加权函数度量每个特征的独立性，距离

越远，特征独立性越高，特征间的冗余度越低。通过计算加权欧氏距离，更准确地衡量特征向量之间的相似性和冗余度，提高对特征向量之间关系的准确度。第 i 个攻击类型特征 x_i 与第 j 个攻击类型特征 y_j 的加权欧氏距离为：

$$NED(\boldsymbol{X},\boldsymbol{Y})=\sqrt{\sum_{K=1}^{N}\frac{w_i}{W}(x_k-y_k)^2} \tag{3}$$

其中，N 为每条记录所拥有的特征个数，即 $N=154$；$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 分别表示数据集中任意两条记录(向量)，x_k 表示 $\boldsymbol{X}$ 中的第 K 个属性，y_k 表示 $\boldsymbol{Y}$ 中的第 K 个属性；$\frac{w_i}{W}$ 表示归一化因子；$w_i=e^{-\frac{|x_i-y_i|}{\sigma}}$，$W=\sum w_i$，$\sigma$ 为调节因子，此处 $\sigma=1$，因此 $w_i=e^{-|x_i-y_i|}$。

综合特征间的皮尔逊相关系数与加权欧氏距离，计算特征与向量之间的关联度：

$$S_{ij}=\frac{\rho_{ij}+NED(\boldsymbol{X},\boldsymbol{Y})_{ij}}{2} \tag{4}$$

关联度取值范围为(0,1)，数值接近 1 时说明两类特征的相关性大且信息的独立性高，可以将其作为有效特征保留；数值接近 0 时说明两种特征不具有相关性且信息冗余程度较高，不建议将其作为有效特征保留。

4 基于多源数据的网络入侵检测模型

通过神经网络构建电力监控系统网络攻击分类模型，用于实现对电力监控系统网络的多源异构数据的分类，具体实现为：

1)卷积神经网络方法：在对网络多源异构数据进行分类前，利用卷积神经网络对特征进行分析，并将输入的多源网络安全数据序列转换为固定长度的特征向量表示。

2)长短期记忆网络方法：根据卷积神经网络提取的特征对电力监控系统网络的多源异构数据进行训练和分类，学习输入序列中的模式和依赖关系，并生成相应的分类结果。

4.1 卷积神经网络层

针对复杂多变的多源网络安全数据的入侵检测数据集，构建卷积神经网络模型，利用一维卷积层处理数据。由于多源网络安全数据通常是时间序列数据或具有序列性质的数据，因此卷积操作能够深入数据内部，高效地捕捉并提炼出数据中隐藏的局部特征，例如某种攻击的网络流量行为、异常的数据包结构等，还可以通过学习大量的样本数据，自动发现新攻击的潜在特征模式，从而提高对新型攻击的检测能力。

卷积神经网络(Convolutional Neural Network，CNN)是一种前馈神经网络，主要由输入层、卷积层、池化层、全连接层和输出层构成。在一维卷积层内部，由 BatchNormalization 层以及 DropOut 层组合形成的一维卷积核沿着输入的多源网络安全数据序列的时间轴方向滑动，计算得到多种与攻击类型相关的局部特征，将多源网络安全数据的特征放入全连接层，在输出层中调用 Softmax 函数进行概率判定，以得到特征处理的卷积网络模型。同时，选择卷积核、表现效果较优的 LeakyRelu 激活函数和 Adam 优化器对模型进行训练。本文提出的一维卷积网络模型如图 2 所示。

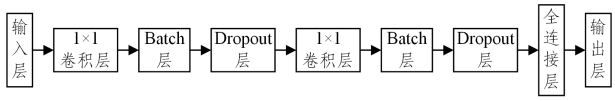

图 2 基于多源网络安全数据的卷积神经网络模型结构

Fig. 2 Structure of convolutional neural network model based on multi-source network security data

4.2 长短期记忆网络层

长短期记忆神经网络(LSTM)是循环神经网络(RNN)的改进变体，LSTM 与 RNN 的整体结构基本相同，但隐藏层设计不同，LSTM 通过记忆单元中的 3 个逻辑门，即输入门、遗忘门、输出门，避免了反向传播过程中的梯度消失和梯度爆炸问题。LSTM 单元结构如图 3 所示。

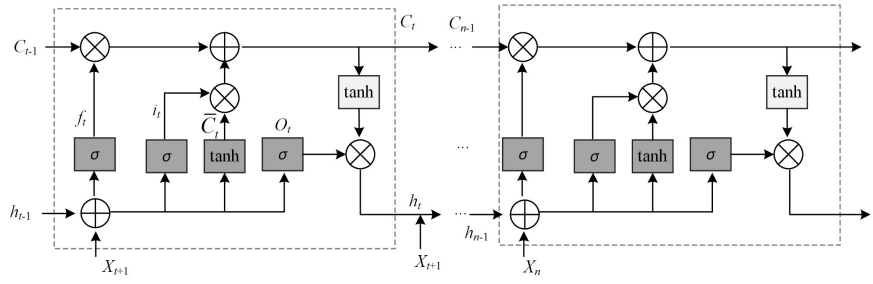

图 3 LSTM 网络内部结构

Fig. 3 Internal structure of LSTM network

图 3 中，C_t 表示记忆单元状态；i_t 表示输入门，用于控制记忆单元更新的信息量，即当前时刻网络的状态有多少信息需要保存到内部状态中；f_t 表示遗忘门，用于控制前一时刻记忆单元状态 C_{t-1} 有多少被保存到记忆单元 c_t 中；o_t 表示输出门，用于控制记忆单元 c_t 输出多少到下一隐藏状态 h_t；h_{t-1} 表示在 $t-1$ 时刻的隐藏层的状态。

LSTM 记忆单元在 t 时刻会接收当前时刻的输入 x_t；$t-1$ 时刻的外部状态 h_{t-1} 和内部状态 C_{t-1}，其中，x_t 和 h_{t-1} 同时作为 3 个门的输入。输入门、遗忘门、输出门的计算方式为：

$$i_t=\sigma(w_i\times x_t+v_i\times h_{t-1}+b_i) \tag{5}$$

$$f_t=\sigma(w_f\times x_t+v_f\times h_{t-1}+b_f) \tag{6}$$

$$o_t=\sigma(w_o\times x_t+v_o\times h_{t-1}+b_o) \tag{7}$$

记忆单元 c_t 通过不同门控数据的记忆和遗忘进行更新，计算方式为：

$$c_t'=\tanh(w_c x_t+v_c h_{t-1}+b_c) \tag{8}$$

$$c_t=c_{t-1}\times f_t+c_t'\times i_t \tag{9}$$

更新后的隐藏状态为：

$$h_t=o_t\otimes\tanh(c_t) \tag{10}$$

其中，ω 为 x 的权值，v 是 h 的权值，b 为偏置项。LSTM 记忆单元接收 x_t 和 h_{t-1}，在激活函数的作用下，产生分别控制输入门、遗忘门、输出门的信号。LSTM 根据当前时刻各门的信号，更新记忆单元 c_t 以及生成当前时刻的状态输出 h_t，并将

其作为后续 $t+1$ 时刻的输入。

4.3 基于神经网络的入侵检测模型

电力监控系统的多源网络数据往往同时具有空间和时间特性，LSTM 专门设计用于处理时间序列数据，能够有效地捕捉多源网络数据中的时间依赖关系，而 CNN 主要处理空间维度的信息，能够通过不同的卷积核和层次结构，适应这种复杂性。同时，CNN 擅长从多源数据中提取丰富的特征，而 LSTM 能够对这些特征进行序列建模。LSTM 通过门函数控制历史数据的记忆和遗忘，适合处理具有时间序列特性的问题。LSTM 能较好地联系上下文，但由于其结构复杂，随着输入的数据增加，计算量也随之增大，从而降低了上下文的联系，使得算法的准确率降低。卷积神经网络通过卷积核的操作可以提取具有事件特点的多源网络安全数据的特征，将特征作为 LSTM 模型的输入以进行电力监控系统网络攻击的检测与分类。利用神经网络进行多源网络安全数据入侵检测的网络结构如图 4 所示。

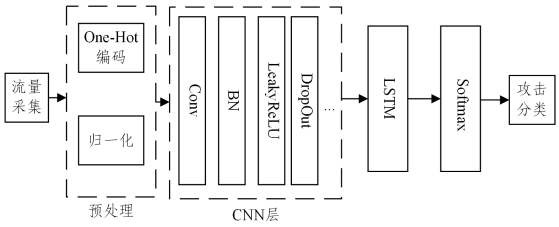

图 4 基于多源网络安全数据的网络结构

Fig. 4 Network structure based on multi-source network security data

多源网络安全数据的电力监控系统网络入侵检测算法步骤如下：

1）基于电力监控系统多源异构数据清洗方法，对构建的训练集与测试集进行数据预处理，对安全数据的特征（包括网络协议、网络服务类型和网络连接状态等数据）进行独热编码。同时，对数据包特征中的连接时间等连续型数值数据进行归一化处理。

2）基于 mRMR 框架的策略，针对多源异构数据实施特征选择，该策略融合了相关系数分析以评估特征与目标变量的紧密程度，并结合多距离加权函数来精细衡量特征间的冗余性。通过对网络流量数据包中的特征进行这一综合筛选过程，旨在精准保留那些既与目标变量高度相关，又保持较低冗余性的有效特征，从而优化后续分析的性能与效率。

3）经过精细的特征选择过程后，将优化后的网络安全数据作为输入，注入到精心设计的卷积神经网络模型中。该模型随后运用其强大的卷积分析能力，深入剖析数据特征，最终输出精确且可靠的分类结果，识别出网络中的潜在威胁或异常行为。

5 实验与结果分析

5.1 评价指标

实验部分使用准确率、精确率、召回率 3 种评价指标对模型的性能进行评估，为了更方便地说明 3 种指标的具体含义，利用 True Positive（TP），False Negative（FN），True Negative（TN）和 False Positive（FP）4 种数据分别表示“实际为正样本，且判断为正样本”“实际为正样本，且判断为负样本”“实际为负样本，且判断为负样本”“实际为负样本，且判断为负样本”“实际为负样本，且判断为正样本”，进而对 3 种评价指标进行计算，具体含义及计算方法为：

准确率（Accuracy，Acc）：表示被正确分类的样本占总样本的比例：

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

精确率（Precision，P）：表示被预测为正的样本中实际为正样本的概率：

$$P=\frac{TP}{TP+FP} \tag{12}$$

召回率（Recall，R）：实际为正的样本中被预测为正样本的概率：

$$R=\frac{TP}{TP+FN} \tag{13}$$

5.2 实验仿真

在神经网络中，模型参数的设定会直接影响模型的检测效率，所以本文针对 LSTM 网络模型中的具体参数进行测试。此外，考虑电力监控系统结构与网络模型的计算轮次、系统计算资源，确定当前仿真环境中的最优参数。

5.2.1 神经网络模型训练批次的大小

神经元数量的多少关系着神经网络学习能力的强弱，一定程度上，神经元的数量越多，学习能力越强。当训练轮次设置为 0～50 时，将具有 32 个、64 个、128 个、256 个、512 个、1 024个和 2 048 个神经元的网络模型进行分类准确率的对比，结果如图 5 所示。考虑到神经网络模型的计算资源与运行时间，确定神经元的数量为 128。

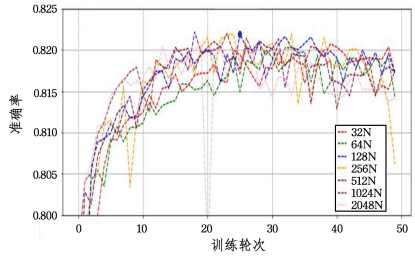

图 5 神经元数量对准确率的影响

Fig. 5 Impact of neuron number on accuracy

5.2.2 神经网络模型全连接层数量

全连接层可以将学到的“分布式特征表示”映射到样本标记空间而实现分类，并且由于全连接层的每一个结点都与上一层的所有结点相连，用于将提取到的与电力监控系统网络结构密切相关的特征综合起来，所以全连接层参数的数量相对较多。考虑到全连接层层数加深后有利于提高模型学习能力，但又存在运算时间增加、效率降低等问题，针对全连接层的层数进行实验，结果如图 6 所示。通过对分类准确度的对比可以确定全连接层层数为 3 时，模型的准确度相对较高，并且可以从一定程度上节省运算时间，提高运算效率。

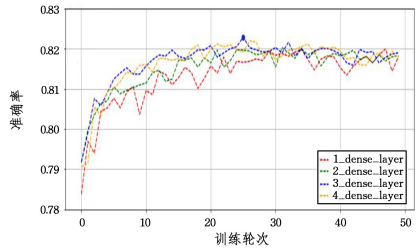

图 6 全连接层数量对准确率的影响

Fig. 6 Impact of number of fully connected layers on accuracy

5.2.3 神经网络模型训练批次的大小

在神经网络模型中，分批训练能够提高模型的适用性，训练批次表明了一个数据块中包含的数据量，并且训练批次的大小会影响模型的优化程度、速度和内存资源的使用情况。考虑到以上问题，对不同训练批次进行比较，结果如图 7 所示。在选择分类准确度较高的批次大小的同时，尽量节省电力监控系统进行网络安全检测时的内存资源，故确定神经网络模型训练批次的大小为 256。

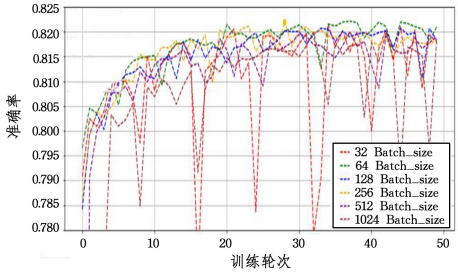

图 7 批次大小对准确率的影响

Fig. 7 Impact of batch size on accuracy

5.2.4 神经网络模型全连接层丢弃率

在全连接层对神经元进行随机丢弃可以使其不参加当前轮次的参数优化，从而防止模型过拟合，提升模型泛化能力，并且每次只有一部分参数更新，可以减慢模型的收敛速度。如图 8 所示，经过对不同丢弃率情况下的模型准确率对比可知，当丢弃率为 20%时，模型的效率相对较高，从而确定全连接层丢弃率为 20%。

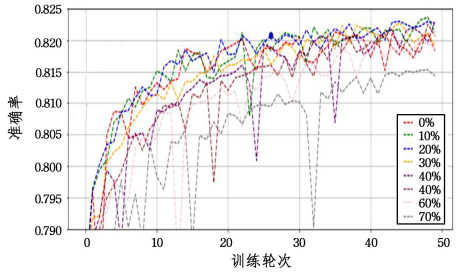

图 8 全连接层丢弃率对准确率的影响

Fig. 8 Impact of fully connected layer dropout rate on accuracy

5.2.5 神经网络模型优化器的选取

神经网络模型的优化器可以更新和计算影响模型训练和模型输出的网络参数，使其接近最优值。如图 9 所示，通过对比 sgd，rmsprop，adagrad，adadelta，adam，adamax 和 nadam 7 种优化器，选取模型准确率最高的优化器——Adam。

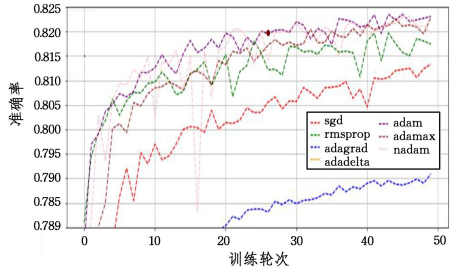

图 9 不同优化器对准确率的影响

Fig. 9 Impact of optimizers on accuracy

5.2.6 神经网络模型学习率的选取

模型学习率的大小会对模型的训练产生很大的影响，学习率数值设置过大，会导致学习难以收敛，学习率数值设置过小，会大大增加训练时间。如图 10 所示，通过对 6 种学习率的对比可确定，当模型学习率设置为 10^{-3} 时，模型的准确率相对较高且可节省模型的训练时间。表 4 列出了 GRU 网络架构的仿真参数设计。

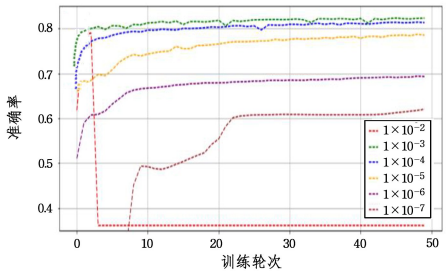

图 10 不同学习率对准确率的影响

Fig. 10 Impact of learning rates on accuracy

表 4 仿真参数设计

Table 4 Simulation parameters design

参数名称	配置
优化器	adam
损失函数	binary_crossentropy
批次	256
训练轮次	25
丢弃率	0.2

采用全特征分类模型与经过上文特征选择算法之后的 KNN，LSTM，GRU，SVM 以及提出的基于神经网络的入侵检测模型进行分类，表 5 为各训练模型的准确率、精确率、召回率对比结果，可知在未使用特征选择算法，即全特征的情况下，所提方法在准确度方面较其他方法有所提高，但由于特征数量过多，该方法的精度与召回率较其他方法相对较低；在使用特征选择算法后，所提方法较其他方法在准确度、精确率、召回率方面均有所提高。

表 5 各模型分类结果

Table 5 Classification results of each model

模型	全特征			特征选择		
	Acc	*P*	*R*	*Acc*	*P*	*R*
KNN	68.50	66.53	67.52	67.52	67.32	67.34
LSTM	60.70	59.47	60.12	67.16	66.07	66.62
GRU	68.50	67.64	68.09	67.68	65.12	66.40
XVM	70.30	69.99	70.15	74.56	73.29	73.92
Ours	70.39	69.00	69.70	78.81	77.10	77.46

使用不同的特征选择方法（所提方法、信息增益法、相关系数法）结合分类模型进行分类的准确率对比如图 11 所示。

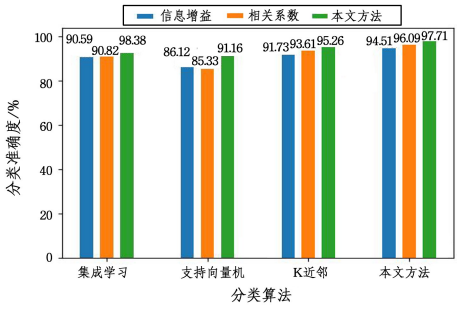

图 11 不同特征选择算法下的分类模型训练结果

Fig. 11 Classification model training results under different feature selection algorithms

由图 11 可以看出，与其他特征选择方法相比，利用所提

的特征选择方法对数据集进行处理后，对后期的分类结果更有利。此外，根据分类训练结果可以看出，所提的入侵检测模型在准确度上比 GRU 模型提高了 11.13%，相比于 SVM 模型提高了 4.25%，相比于 KNN 模型提高了 11.29%。所提的入侵检测模型在各模型准确性中表现最优。

结束语 为有效检测电力监控系统网络异常行为，利用样本基数与样本数据具有时间属性的特点，提出了一种电力监控系统多源数据清洗方法及入侵检测方法，该方法结合神经网络算法与基本分类方法检测网络的异常行为，有效解决了数据集特征属性过多导致后期分类准确度下降等问题。在网络公开入侵检测数据集上的结果表明该算法在检测准确度、精确率、召回率等方面均有所提高，取得了良好的检测效果，但所提方法在电力监控系统的实际验证、入侵类型的检测范围、入侵对系统的影响程度分析等方面还需进行深入研究。

参 考 文 献

[1] SONG L, FAN Y, LIU M, et al. State estimation method of a new energy power system based on SC-DNN and multi-source data fusion[J]. Power System Protection and Control, 2023, 51: 177-187.

[2] WU B, HU Y. Analysis of substation joint safety control system and model based on multi-source heterogeneous data fusion[J]. IEEE Access, 2023, 11: 35281-35297.

[3] SAHU A, MAO Z, WLAZLO P, et al. Multi-source data fusion for cyberattack detection in power systems[J]. arXiv: 2101.06897.

[4] GUO H, LIU Y, XU L, et al. Recognition of demand response potential resident users based on multi-source heterogeneous data association rule analysis[J]. Dianwang Jishu/Power System Technology, 2023, 47(5): 1950-1960.

[5] HE J, JIANG W, LIU G, et al. Advanced monitoring infrastructure system architecture for massive multi-source data[C]// 2023 IEEE 13th International Conference on Electronics Information and Emergency Communication (ICEIEC). IEEE, 2023: 148-152.

[6] JIANG L, WANG W, ZHENG S, et al. A novel data edge mining algorithm with euclidean distance weighted optimization for integrated energy station[C]// 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2023, 7: 67-71.

[7] PARK J K, LEE H, KIM W, et al. Degradation feature extraction method for prognostics of an extruder screw using multi-source monitoring data[J]. Sensors, 2023, 23(2): 637.

[8] LI Y, YU X. The multi-mode FBN multi-source heterogeneous data fusion technology based on AES algorithm[C]// 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, 2023: 1198-1202.

[9] DONG Y, WEN C, WANG Z. A motor bearing fault diagnosis method based on multi-source data and one-dimensional lightweight convolution neural network[C]// Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering. 2023: 272-283.

[10] XIONG Z, LI W, CAI Z. Federated generative model on multi-source heterogeneous data in iot[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 202310537-10545.

[11] ANKITDESHPANDE Y, KARTHI R. Development of intrusion detection system using deep learning for classifying attacks in power systems[C]// Soft Computing: Theories and Applications (SoCTA 2019). Singapore: Springer, 2020: 755-766.

[12] HAN Y, WANG Y, CAO Y, et al. A novel wrapped feature selection framework for developing power system intrusion detection based on machine learning methods[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(11): 7066-7076.

[13] DASGUPTA R, PRAMANIK M, MITRA P, et al. Intrusion detection for power grid: a review[J]. International Journal of Information Security, 2024, 23(2): 1317-1329.

[14] LIU J Q, WANG R, CAO J W. An overview of new information transmission methods for power systems[J]. Journal of Northeast Dianli University, 2024, 44(4): 1-8, 76.

[15] ZHANG D, ZHANG Y, ZANG X X. Anomalous Intrusion Detection Method for Surveillance Video Based on Self-Organising Mathematical Models[J]. Journal of Northeast Dianli University, 2022, 42(4): 63-69.

[16] DURAIRA D, VENKATASAMY T K, MEHBODNIYA A, et al. Intrusion detection and mitigation of attacks in microgrid using enhanced deep belief network[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2024, 46(1): 1519-1541.

[17] RAJASOUNDARAN S, KUMAR S V N S, SELVI M, et al. Secure and optimized intrusion detection scheme using LSTM-MAC principles for underwater wireless sensor networks[J]. Wireless Networks, 2024, 30(1): 209-231.

[18] NUGROHO S A, TAHA A F, QI J J. Robust dynamic state estimation of synchronous machines with asymptotic state estimation error performance guarantees[J]. IEEE Transactions on Power Systems, 2020, 35(3): 1923-1935.

JIANG Yakun, born in 1970, master of engineering. His main research interest is dispatching automation of power system network security.