基于胶囊网络的工业互联网入侵检测方法

胡向东,李之涵
(重庆邮电大学自动化学院/工业互联网学院,重庆 400065)

摘 要: 工业互联网在快速发展的同时,面临着严峻的信息安全风险.针对传统入侵检测方法准确性低、难以适应工业互联网海量不平衡数据的问题,提出一种基于胶囊网络的工业互联网入侵检测方法.首先,基于残差块构建特征提取模块,引入全局平均池化层得到高质量的数据特征;其次,使用动态路由算法,通过迭代的方式对入侵数据特征进行聚类,在胶囊网络模块完成数据分类.基于Modbus/TCP协议的气体管道传感器网络数据集的测试结果表明,该方法可以在隐性提取特征的同时改善检测准确率.与所列算法对比,本文方法提高了检测指标,对不平衡数据有更强的鲁棒性,更接近工业互联网入侵检测技术需求.

关键词: 工业互联网;入侵检测;胶囊网络;残差网络
中图分类号: TN918.91 **文献标识码:** A **文章编号:** 0372-2112(2022)06-1457-09
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20201275

Intrusion Detection Method Based on Capsule Network for Industrial Internet

HU Xiang-dong, LI Zhi-han
(*School of Automation/School of Industrial Internet, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Industrial internet is rapidly growing up while encountering severe information security risks at the same time. Aiming at the problem that traditional intrusion detection methods are low in accuracy and difficult to adapt to the massive unbalanced data of industrial Internet, an industrial Internet intrusion detection method based on capsule network is proposed. Firstly, a module involved feature extraction module is constructed based on the residual block, and a global average pooling layer is introduced to get high-quality data features. Secondly, the dynamic routing algorithm is introduced. The intrusion data features are clustered through iteration, and classification are completed in the module based on capsule network. The test results out of the data set from sensor network with Modbus/TCP protocol used in gas pipeline show that the method can improve the accuracy rate while extracting features implicitly. Compared to the listed algorithms, the proposed method in this paper performs better in test indexes with stronger robustness to unbalanced data and is closer to meet the needs of intrusion detection from industrial Internet.

Key words: industrial Internet; intrusion detection; capsule network; residual network

1 引言

近年来,工业互联网在智能制造、燃气供给、电力水利等众多工业领域兴起[1],成为推动经济发展和新一代基础设施建设不可或缺的新兴产业支撑性技术.工业控制网络原本相对封闭,与外界物理隔离,网络建设本身更多关注运行稳定性和功能安全,缺乏对网络开放条件下信息安全问题的全面设计.随着工业信息化深度发展,针对工业互联网的攻击日益频繁,破坏性逐渐增大[2].典型地,“震网”病毒通过西门子设备漏洞对伊朗核电站的攻击为工业互联网信息安全敲响了警钟.尽管传统互联网中安全产品基于历史积累和迭代达到了一定成熟度,但限于工业互联网自身的资源特点、运行模式和网络属性等,并不能将现有方法直接移植入工业互联网,往往需量身定制针对性解决方案[3].如通过对工业系统通信模式和加密模式的研究可知,工业网络与家庭和办公室网络并不兼容,因此普通入侵检测系统(Intrusion Detection System,IDS)无法直接

收稿日期:2020-11-12;修回日期:2021-04-21;责任编辑:王天慧
基金项目:教育部-中国移动科研基金(No.MCM20150202)

适应工业应用[4]. 同时,因为相对独立,工业互联网中攻击行为的数量远低于传统互联网,这种情况也会给入侵检测带来困难.

与此同时,构建IDS的方法呈现出多样化发展趋势,新思路、新方法不断涌现. 文献[5]使用熵离散化算法和决策树构建分类器对多个应用进行分类,而后对其进行了稀疏化处理. 文献[6]使用OpenPLC平台以及AES-256加密模拟数据采集与监视控制(Supervisory Control and Data Acquisition,SCADA)系统,在此基础上使用无监督的k-means算法对代码注入攻击、拒绝服务攻击和拦截进行了检测. 文献[7]利用感知哈希矩阵量化属性,而后使用K近邻投票原则完成入侵检测任务. 文献[5~7]通过人工选取和组合特征的手段完成分类器的构建,普遍存在检测准确率偏低、系统鲁棒性差等问题,特征选取的优劣也极大影响了实验效果.

在浅层学习方法发展的同时,基于深度学习技术的IDS也取得了巨大进步. 文献[8]设计了使用支持向量机和深度信念网络的混合IDS用于工业控制系统,但该文献使用了NSL-KDD数据集进行仿真,NSL-KDD较为老旧,且并不适用于工业控制系统环境. 文献[9]利用基于离差标准化(Min-Max Normalization,MMN)的卷积神经网络对校园网流量进行分析,该模型开销较小且易于训练,很好地解决了参数选取问题. 但该分类方法结构与LeNet-5差异较小,结构简单,难以应对海量复杂数据的特征学习. 文献[10]针对入侵样本分布不均衡问题,使用基于窗口的实例选择算法清洗训练集并基于循环神经网络构建了入侵分类模型. 虽然取得了较高的准确率,但实验使用了复杂的预处理过程,难以体现深度学习隐性提取特征的优势. 文献[11]使用条件深度信念网络检测智能电网中的攻击,该文献使用总线测试系统进行仿真,并提供了与人工神经网络和支持向量机等方法的对比. 这些方法从多角度探讨了IDS的不同方案,但部分方法使用复杂手段处理数据,难以解决不同分类算法对特征选取的差异性要求. 多数模型结构单一,存在模型收敛慢、样本分布不均匀时鲁棒性差等问题.

2017年,文献[12]提出基于向量的胶囊网络(Capsule Network,CapsNet). 该网络引入了向量胶囊层和动态路由算法,用胶囊表示神经元的集合,使用动态路由的方法连接不同隐藏层之间的胶囊以映射不同特征间的相对关系. 胶囊网络改善了传统卷积神经网络对目标位置不敏感的问题. 例如,文献[13]关于高光谱图片的分类研究中,即使测试样本数量远大于训练样本数量,胶囊网络依然取得了较好的分类成绩. 但因动态路由算法无法分享各神经元的权重,胶囊网络的参数量远大于传统的卷积神经网络.

通过对上述文献的分析可知,深度学习有可能从工业互联网数据中提取优质特征,从而创建更好的模型. 本文受胶囊网络启发,在此基础上引入残差结构对其进行改进,构建融合残差块的胶囊网络(Residual Capsule Network,RCN)对工业互联网数据关联特征进行学习,构建入侵检测模型实现对网络流量的有效处理. 该方法避免了人工特征提取的复杂流程,很好地提高了入侵样本分布不均衡背景下的检测精度,缩短了模型训练时间.

2 融合残差块的胶囊网络

文献[12]提出的胶囊网络是一个浅层神经网络. 其结构如图1所示.

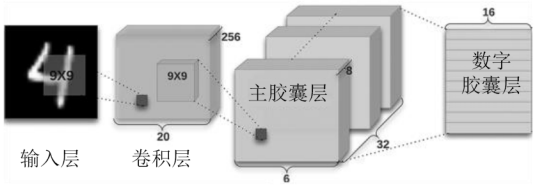

图1 胶囊网络基本结构

胶囊网络引入动态路由算法对向量胶囊之间的关系进行迭代. 因参数量多于传统卷积神经网络,在多数节点算力有限的场景下,胶囊网络难以在工业互联网中分布实施. 同时,工业互联网数据特性又使得传统模型性能欠佳. 为了降低胶囊网络的计算开销,提高识别精度,本文通过引入残差结构的方式改进胶囊网络,其结构如图2所示.

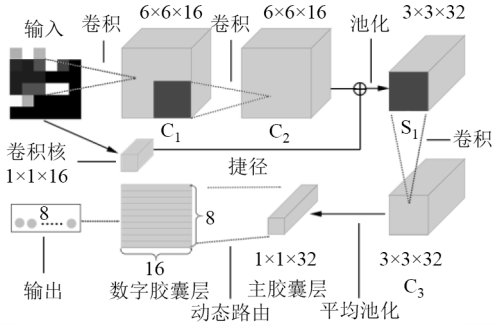

图2 融合残差块的胶囊网络结构

RCN使用残差结构对流量特征进行降维以提高胶囊输入质量,令胶囊的预测更加快速精准. 融合残差块的胶囊网络包括由残差块、卷积层和池化层构成的残差网络模块以及由主胶囊层和数字胶囊层构成的胶囊网络模块.

2.1 残差网络模块

微软研究院的He[14]等人于2016年提出了深度残差神经网络(Deep Residual Network,DRN). DRN又名

ResNet,通过引入残差块的方式完成了多达152层网络的构建,极大地改善了因网络深度增加导致的性能退化问题. 同时,DRN在分类步骤前引入一个全局平均池化层提取每个特征图的均值,可以对数据进行降维的同时避免出现过拟合. 本文在DRN结构的基础上构建特征提取的残差网络模块,从而提高胶囊网络的输入特征质量.

2.1.1　卷积层

卷积层是对图像进行抽象的主要结构. 卷积层由数个各异的特征图组成,卷积层对上层的一小块区域进行卷积操作,形成下一层的节点[15],这块区域被称为卷积核或滤波器. 卷积的具体形式为

$$x_j^l=f\left(\sum_{i\in M_j}x_i^{l-1}k_{ij}^l+b_j^l\right) \tag{1}$$

式(1)中,x_j^l表示卷积操作之后第l层中第j个位置输入;x_i^{l-1}表示第$l-1$层中的第i个位置输入;k_{ij}^l表示连接第$l-1$层和第l层的卷积核;b_j^l表示第l层中第j个位置的偏置. 本文在RCN的C_1和C_2层使用16个步长为1的卷积核对输入进行卷积处理,并通过填充0保持卷积前后特征图的尺寸不变.

因为卷积运算不包含非线性成分,所以经过式(1)计算之后得到的输出数据还需要使用激活函数进行非线性处理,使网络具有拟合复杂特征的能力. 在卷积神经网络中,ReLU函数可以大幅提高网络稀疏性,提高模型效率[16]. 本文采用ReLU作为各层的激活函数.

2.1.2　残差块

残差块是残差网络的主要构成部分,有不同的表示形式,共同特点为引入了捷径,把输入传到输出作为结果的一部分. 一种典型的残差块结构如图3所示.

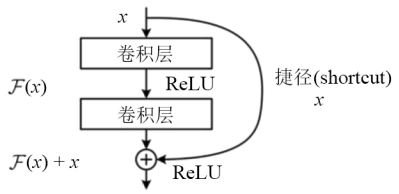

图3　残差块结构

令x为输入,y为输出,$\boldsymbol{W}_i$为权重矩阵,$\mathcal{F}\left(x,\left\{\boldsymbol{W}_i\right\}\right)$为待学习残差函数,残差块可表示为

$$y=\mathcal{F}\left(x,\left\{\boldsymbol{W}_i\right\}\right)+x \tag{2}$$

若x的维度与残差函数的维度不同,为了实现输入与输出的融合,利用线性投影矩阵$\boldsymbol{W}_s$改变维度. 具体公式为

$$y=\mathcal{F}\left(x,\left\{\boldsymbol{W}_i\right\}\right)+\boldsymbol{W}_s x \tag{3}$$

为实现维度转换,本文在捷径中使用一组尺寸为1×1×16的卷积核将输入转化为合适的形式.

2.1.3　池化层

池化层又称采样层,一般应用在卷积层之后,主要作用是对数据进行降维,同时对上层特征图进行压缩形成新的特征图. 这种处理方式既降低了网络的复杂度,又有效保留了原图像的主要特征信息. 池化层可以表达为

$$x_j^l=f\left(\beta_j^l\,\mathrm{down}\left(x_j^{l-1}\right)+b_j^l\right) \tag{4}$$

式(4)中,down(·)为次抽样函数,通常对输入特征图局部进行加权求和;β为按需设置的乘性参数;b为偏置.

本文使用最大池化和全局平均池化2种方法,其主要区别在于使用的次抽样函数不同. 在残差网络模块的S_1层使用步长为2,采样核尺寸为2×2的最大池化层进行数据降维,而后对C_3层输出进行全局平均池化处理,得到一组32维的张量作为胶囊网络模块的输入.

2.2　胶囊网络模块

传统的卷积神经网络通过卷积核抽象图像特征,使用全连接层输出分类结果. 其标量化的计算方式对物体空间关系辨识度差. 胶囊网络使用向量化的神经元(即胶囊)来代替标量神经元节点,改变了传统神经网络标量与标量相链接的架构. 每个胶囊携带的信息从一维增加到多维,向量的方向代表了图像中出现的特定实体的多种属性,例如大小、相对位置、纹路等,向量的长度则表示不同属性的存在概率. 为了匹配胶囊间反向传播需求,文献[12]针对胶囊间迭代关系提出动态路由算法,其核心思想是胶囊的权重由低层次胶囊输入与高层次胶囊输出的相似度决定. 如果低层胶囊的输入与高层胶囊的输出具有较高相似度,则这些低级别胶囊的路由即为较高级别的胶囊.

首先,高层胶囊由低层胶囊计算得出. 动态路由初始阶段,L层的第i个胶囊连接到$L+1$层的胶囊j的概率公式为

$$c_{ij}=\frac{\exp\left(b_{ij}\right)}{\sum_j\exp\left(b_{ij}\right)} \tag{5}$$

式(5)中,b_{ij}是胶囊i连接到胶囊j的先验概率. 在路由更新时,首先计算L层胶囊i对$L+1$层胶囊j的输出的预测胶囊$u_{j|i}$,即

$$u_{j|i}=\boldsymbol{W}_{ij}\times u_i \tag{6}$$

式(6)中,$\boldsymbol{W}_{ij}$为转换矩阵,u_i为L层胶囊i. 在计算预测胶囊之后,通过式(7)和式(8)计算高层胶囊,计算v_j的过程可以用挤压函数表示,即

$$s_j=\sum_j c_{ij}\times u_{j|i} \tag{7}$$

$$v_j=\frac{\left\|s_j\right\|^2}{1+\left\|s_j\right\|^2}\frac{s_j}{\left\|s_j\right\|}=\text{squash}\left(s_j\right) \tag{8}$$

式(8)中,s_j为L层胶囊的总输入,v_j为L+1层胶囊j的输出.将v_j和预测胶囊$u_{j|i}$用于更新b_{ij},从式(5)开始新一轮的循环.动态路由算法迭代过程如算法1所示.

算法1 动态路由算法

输入:迭代次数n,胶囊层数l

过程:

1.初始化$b_{ij}=0$

2.**FOR** n **DO**

3. $c_{ij}=\dfrac{\exp\left(b_{ij}\right)}{\sum_j\exp\left(b_{ij}\right)}$

4. $s_j=\sum_j c_{ij}\times u_{j|i}$

5. $v_j=\text{squash}\left(s_j\right)$

6. $b_{ij}\leftarrow b_{ij}+u_{j|i}\cdot v_j$

7.**END FOR**

输出:v_j

2.3 损失函数和优化算法

胶囊用向量长度来表示其表征内容出现的概率,输出的概率总和并不等于1.所以不同于传统分类任务常用的交叉熵损失,本文采用间隔损失构建网络的损失函数.间隔损失函数可表示为

$$\begin{aligned}L_c=&T_c\max\left(0,m^+-\left\|v_c\right\|\right)^2\\&+\lambda\left(1-T_c\right)\max\left(0,\left\|v_c\right\|-m^-\right)^2\end{aligned} \tag{9}$$

式(9)中,c表示类别;T_c表示第c类入侵是否存在;v_c表示在输出层胶囊的长度,即样本属于第c类的概率;m^+为惩罚假阳性的上界,m^-为惩罚假阴性的下界;λ为比例相关系数,用于调整两者比重.本文分别设置λ,m^+和m^-的值为0.25,0.9和0.1.

动态路由算法解决了胶囊层之间的权重更新问题.但动态路由仅存在于胶囊之间,为了提高网络的收敛能力,还需引入反向传播过程.本文使用Adam方法作为损失函数优化算法,通过迭代最小化损失值以更新神经元权重,令RCN平稳收敛.

2.4 入侵检测模型架构

在RCN的基础上,本文以工业互联网为对象,提出如图4所示的入侵检测模型.

该模型通过预处理强化网络数据映射图像的相对关系,充分挖掘数据信息特征,以RCN为核心执行入侵检测任务,主要包括以下几个模块.

(1)数据预处理模块:将工业互联网数据进行标准化、归一化处理并映射为灰度矩阵.而后,将其转换为灰度图像以便于观察和处理.

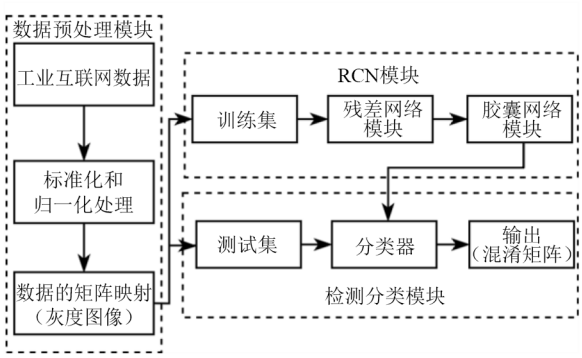

图4 基于RCN网络的入侵检测模型架构

(2)RCN模块:将灰度图像作为模块输入,通过残差网络结构提取数据特征,经过挤压函数处理后送入胶囊模块对特征进行聚合,使用动态路由算法更新胶囊间权重.

(3)检测分类模块:依据数据类别训练分类器,使用分类器对预处理后的攻击样本进行检测,输出多维混淆矩阵,通过混淆矩阵可观察检测结果.

3 数据准备和预处理

3.1 网络特点与数据分析

3.1.1 约束入侵检测方案构建的工业互联网特点

现阶段,工业互联网主要由工厂内部网络与外部网络2部分组成.外部网络用于连接工厂和客户、工厂和供应链等,主要由公共互联网承担.内部网络主要用于工业控制,由工控网络发展而来,主要包括工业生产数据采集与监视控制系统、分布式控制系统、可编程逻辑控制器以及内部生产运营决策支持系统等.

在现代智能工厂等较为复杂的工业系统中,其控制器通常分布式部署,网络包括多个实体和大量传感器连接.工控网络中的各种设备通常由不同的制造商供应,使用区别于传统互联网的特定协议.其网络节点计算能力差异大,实时性要求高,这些因素使工业互联网与民用网络存在较大不同[17,18].

此外,工业互联网流量特征在于流量的规则性和协议特殊性,具有稳定的吞吐量和周期模式,有清晰的数据包和可预测的数据流向[19].这种特点可以进行监督学习,适合基于异常的入侵检测技术的开发和实施.同时,由于工业系统的稳定性和相对隔离性,每天产生的数据是海量的,但网络攻击占比较低,在统计上表现为数据的严重不平衡.

3.1.2 实验数据分析

基于对实验数据的典型性、广泛接受性、系统性和研究对象适宜性等要求,本文采用美国密西西比州立大学SCADA实验室采集的气体管道数据集(gas pipe-

line)[20]对检测模型进行验证. 该工业互联网数据集来源于实验室规模应用Modbus/TCP协议的气体管道平台. 该平台包括压缩机、压力表等传感器和使用电磁阀控制小型气密管道,使用比例积分微分控制方案维持管道气压. 通过基于RS-232的网络数据记录器监视和存储Modbus流量.

该平台使用线路插件捕获数据日志并进行攻击注入,通过在VMware虚拟机上运行的C程序监视串行端口的通信,为流量标记时间戳并记录在日志文件中. 以命令注入攻击为例,向平台发送恶意命令会尝试开关压缩机或调整安全阀的状态,通过记录网络流量特征、过程控制和传感器状态可形成该类攻击的具体数据. 气体管道数据集将数据分为注入攻击、拒绝服务攻击、侦查攻击和正常数据4个大类,其中,注入攻击又可以分为5个子类. 数据的具体类别及分布情况如表1所示.

表1　气体管道数据集样本类别分布

数据类型	数量	标签值	描述
Normal	61 156	0	正常数据
NMRI	2 763	1	简单恶意响应注入攻击
CMRI	15 466	2	复杂恶意响应注入攻击
MSCI	782	3	恶意状态命令注入攻击
MPCI	7 637	4	恶意参数命令注入攻击
MFCI	573	5	恶意功能命令注入攻击
Dos	1 837	6	拒绝服务攻击
RECO	6 805	7	侦查攻击

由于多种原因,例如不同的昼夜模式、数据缺少相关性以及分布的不同,用于描述公共互联网流量的现有模型不能直接应用于工业互联网. 与有20多年历史的KDD99[21]以及其他数据集相比,gas pipeline更符合工业互联网现状,其数据构建方式更符合真实网络环境要求.

为了充分验证RCN在工业互联网入侵检测领域的有效性,采用随机的方式抽取气体管道数据集50%的数据用于模型训练和测试,并按4∶1的比例随机划分训练集与测试集,二者分布如表2所示.

由表2可知,训练集和测试集均具有明显的数据不平衡性. 数据不平衡引起的分类问题始于二分类中的数据偏态,会使检测器对多数类样本产生偏倚[22],当少数类样本在一些情况下至关重要时,不能区分少数类样本会使检测失去实际意义.

以训练集为例,正常数据占数据总量比例高达63%,而数据量最少的恶意功能命令注入攻击MFCI仅占数据总量的0.57%,相较于多数类样本所占的比重极小,造成少数类样本的相对稀缺.

表2　气体管道数据集样本类别分布

数据种类	数据来源	
	训练集	测试集
Normal	24 513	6 107
NMRI	1 107	268
CMRI	6 036	1 622
MSCI	346	71
MPCI	3 099	746
MFCI	222	56
Dos	776	169
RECO	2 709	663
合计	38 808	9 702

3.2　网络特点与数据分析

3.2.1　数据的标准化和归一化

气体管道数据样本中数值差别大,异常值、离群值数量多,会对入侵检测模型的收敛速度和精度造成负面影响,因此需要对数据依次进行标准化和归一化处理. 气体管道数据共有27类属性实例,包括26类数据特征和一类标签属性. 将标签分离,首先假设数据集可以用n行m列的矩阵$\boldsymbol{T}$构成,即

$$\boldsymbol{T}=\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nm} \end{bmatrix} \tag{10}$$

令$\boldsymbol{A}_i$为一类特征,可将$\boldsymbol{A}_i$表示为

$$\boldsymbol{A}_i=\left[B_{1i},B_{2i},\cdots,B_{ni}\right]^{\mathrm{T}} \tag{11}$$

那么矩阵$\boldsymbol{T}$可以表示为

$$\boldsymbol{T}=\left[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_m\right] \tag{12}$$

计算每类特征数据的标准差(Standard Deviation, SD). 若SD≥8,则需要对这类特征数据进行标准化处理,令$\boldsymbol{A}_i'$为处理后的数据,$\boldsymbol{A}_i$为需处理的原始数据,标准化过程可以表示为

$$\boldsymbol{A}_i'=\arctan\boldsymbol{A}_i \tag{13}$$

完成标准化处理后的数据需进一步归一化. 若SD<8,则跳过标准化过程直接进行归一化处理,方法如式(14)所示:

$$\boldsymbol{A}_i''=\frac{\boldsymbol{A}_i'-\min\left\{B_{1i},B_{2i},\cdots,B_{ni}\right\}}{\max\left\{B_{1i},B_{2i},\cdots,B_{ni}\right\}-\min\left\{B_{1i},B_{2i},\cdots,B_{ni}\right\}} \tag{14}$$

3.2.2　矩阵映射和可视化处理

完成标准化和归一化处理后得到范围在[0,1]之间的数据集. 为构建合适的输入形式,使用一个数值代表一个矩阵中的一个灰度像素点,将数值乘以255后,将26位数据特征填充到6×6大小的灰度矩阵中. 因矩阵维数大于数据特征位数,需要在矩阵末尾进行补零

操作.

从各类数据中各随机选取一组,将其映射的灰度矩阵转化为图像,得到如图5所示的图片集.其中,数据值越小,对应矩阵位置越接近黑色,反之亦然.可以看出,不同种类数据所映射的图片之间有较为明显的区别,而同种类的数据映射的图片有一定的相似性.从可视化处理的结果看,预期使用RCN学习特征可以取得较好的效果.

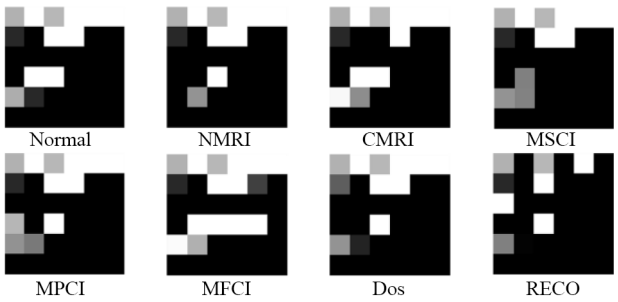

图5 不同数据种类的可视化表示

4 实验与结果分析

4.1 实验环境与超参数设置

为模拟工业互联网环境,使用一台工控计算机训练网络模型,训练过程不使用图形加速卡.实验的软硬件环境配置如表3所示.

表3 实验环境配置

类别	参数
操作系统	Centos 7
处理器	Intel Core i7-8550U
内存	2×4 GB DDR4 2133 MHz
Keras	2.2.4
Tensorflow	1.14
Python	3.6.10

进行参数组合训练,依据测试结果确定模型训练的超参数设置如下:每次迭代训练从训练集数据中选取的数据量为256,动态路由迭代次数为3.

4.2 评价指标

入侵检测算法典型的评价指标包括准确率(Acc)、误报率(False Negative Rate,FNR)和漏报率(False Positive Rate,FPR),但因工业互联网入侵数据不平衡性严重,占总数量绝大多数的正常数据会使传统指标发生偏移.为保证评估的全面性,本文综合采用准确率、漏报率、误报率和F1值作为评价指标,其中,F1值由查全率(Recall)和查准率(Precision)定义.指标可由式(15)~式(20)定义:

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN}} \tag{15}$$

$$\mathrm{FNR}=\frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} \tag{16}$$

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}} \tag{17}$$

$$\mathrm{F1}=\frac{2\ \mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{18}$$

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{19}$$

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{20}$$

式(15)~式(20)中,TP代表真正类,表示本属于攻击的样本被正确预测为攻击的样本数;FN代表假负类,表示将攻击误报为正常样本的数目;FP代表假正类,表示本属于正常的样本被错误预测为攻击的样本数;TN代表真负类,表示本属于正常的样本被准确预测为正常的样本数.

4.3 实验结果分析

为了评估RCN在工业互联网入侵检测中的指标,除CapsNet外,本文选取深度学习领域的BiLSTM,GRU,MMN-CNN[9]和传统机器学习方法PSO-SVM进行对比实验.

为评估模型训练情况,BiLSTM,GRU,MMN-CNN使用交叉熵作为损失函数,CapsNet和RCN使用间隔损失函数.为公平起见,所有模型使用预处理后的数据,并针对输入形式做相应变换.

4.3.1 数据预处理分析

气体管道数据集离群值多,以气压值属性为例,其最小值为-6.81×10^{37},最大值为6.15×10^{36}.如果直接进行归一化处理,那么这些极值很大的离群值会使其余数据过于集中,无法反映数值的相对大小.本文在预处理过程中使用反正切函数对部分数据进行标准化处理,减小了数据的整体离散性.

图6为一个NMRI类攻击数据进行标准化处理前后对比图.可以看出,标准化处理后的可视化图像灰度特征变化显著,有利于进行深度特征提取.为进一步验证标准化处理的效果,保持其他步骤不变,使用RCN和GRU对预处理前后的数据进行实验,结果如图7所示.可以直观地看出,预处理能有效提升对入侵数据的检测效果.

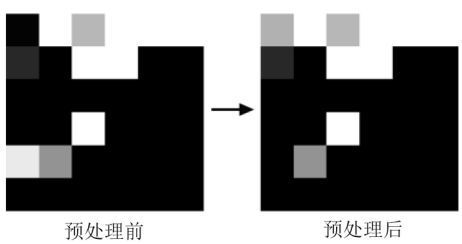

图6 预处理前后对比图

4.3.2 模型收敛性分析

上一节对数据预处理效果进行了分析,图8为不同模型的损失值随迭代次数的变化情况.因为RCN和

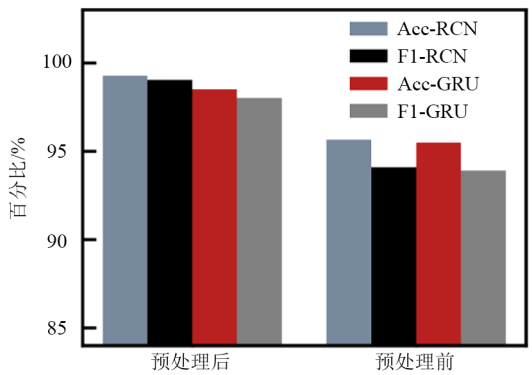

图7　预处理前后指标值变化

CapsNet采取了间隔损失作为评估函数,所以对其损失值进行单独比较.

从图8中可以观察出RCN的训练损失曲线平稳较快收敛. 结合图9,可以直观地看出所有检测模型都有一定程度的检测能力,但本文方法在收敛速度与精度上均体现出显著的优势. BiLSTM和GRU容易陷入局部最优,曲线波动较大. MMN-CNN无法对图像相对位置关系进行处理,训练精度整体较低. 在引入残差网络进行改进后,可以看到RCN的精度曲线比CapNet更加平稳,这说明残差结构的引入显著提高了胶囊层分类质量.

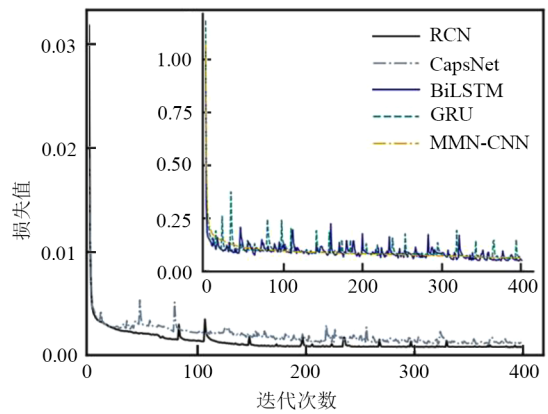

图8　损失值随迭代次数变化的曲线

4.3.3　检测指标比较

为从整体上比较各模型的预测能力,训练各个模型并进行分类测试. 因为深度学习权重初始化具有随机性,为保证数据的可靠性,对所有模型进行多次训练并取平均值,实验结果如表4所示. 和其他检测模型相比,RCN的4项指标均为最高,F1值达到99.03%,相较于GRU和CapsNet分别高出1.03%和1.48%,这说明残差网络模块显著提高了胶囊层分类质量,充分提取了入侵数据特征. BiLSTM和GRU可提取时序特征,但漏报率和误报率相对较高,说明在训练过程中丢失数据特征信息较多. MMN-CNN和RCN同属于卷积神经网络,但准确率低于RCN,主要原因为胶囊网络使用向量胶囊作为神经元,保留了丰富的图像信息. PSO-SVM各项指标最差,主要原因是支持向量机适用于解决小批量样本的线性回归问题,不能满足IDS对海量数据处理的指标要求.

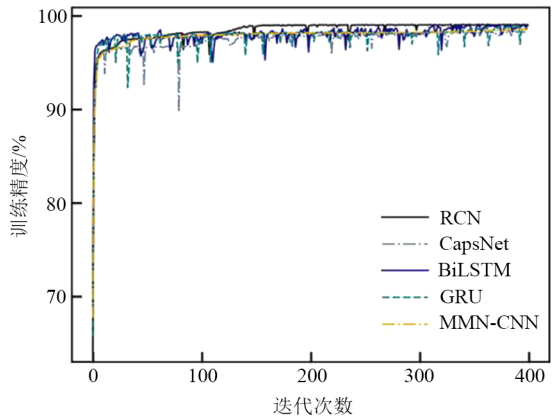

图9　训练精度随迭代次数变化曲线

表4　不同算法下的检测结果

模型	准确率/%	漏报率/%	误报率/%	F1/%
CapsNet	98.18	1.84	1.82	97.55
BiLSTM	98.30	1.56	1.78	97.72
GRU	98.51	1.17	1.69	98.00
MMN-CNN	98.48	1.11	1.75	97.98
PSO-SVM	96.66	5.42	2.11	95.45
RCN	99.28	1.08	0.51	99.03

使用充分训练的模型进行分类测试,最终输出如图10所示的8维混淆矩阵.

混淆矩阵显示了各类测试样本分类后的结果,其

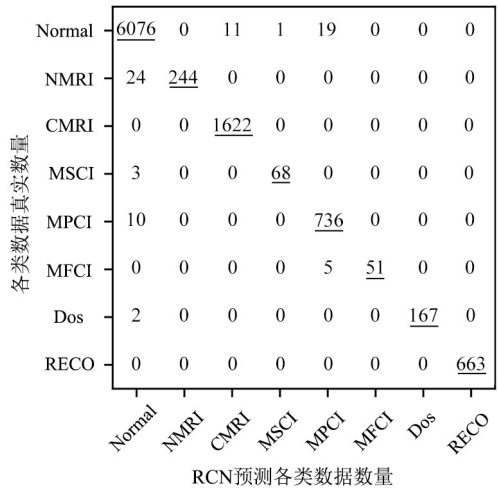

图10　混淆矩阵

中,用下划线标识的数字表示各类数据被正确预测的数量. 将各类数据进行整理,得到RCN对不同攻击类别的检测准确率,如表5所示.

表5 不同攻击类别的检测准确率

NMRI	CMRI	MSCI	MPCI	MFCI	Dos	RECO
91.0%	100%	95.8%	98.7%	91.1%	98.8%	100%

结合F1分数,可以看出,RCN在没有进行数据增强的情况下,对样本数据量较小的各类攻击取得了优秀的检测效果,这说明RCN对图像特征进行了合理聚合,有效地降低了数据不平衡带来的不利影响,泛化性很强.

RCN虽然能达到预期检测效果,但因动态路由算法较为复杂,仍存在改进空间. 经测试,在本文算法中,动态路由算法耗费了总训练时间的41%,限制了RCN模型的识别效率.

4.3.4 运行时间分析

本文所指运行时间包括2部分:模型训练时间和模型预测时间. 模型的运行时间与模型的复杂度和训练迭代次数相关. 通过实验记录了各网络模型的运行时间,结果如表6所示.

RCN与CapsNet产生的训练时间差主要来源于动态路由算法的参数量. CapsNet直接将大维度数据输入主胶囊层,极大地增加了训练时间. GRU具有2个门结构,相对于BiLSTM减少了参数量. MMN-CNN检测速度在所列深度学习方法中最快. 这3种网络运行时间较短,但结合表4综合考虑,总体检测效果差于RCN. PSO-SVM由于结构简单,执行时间代价较低. 但其检测指标最差,预处理过程需要人工进行特征筛选,且未计入运行时间,难以满足入侵检测系统智能化的发展趋势.

表6 不同模型的运行时间对比

数据种类	时间/s		
	训练	预测	总和
CapsNet	1446.57	0.79	1447.36
BiLSTM	391.67	0.42	392.09
GRU	296.52	0.37	296.89
MMN-CNN	161.80	0.23	162.03
PSO-SVM	1.12	0.07	1.19
RCN	517.11	0.56	517.67

5 结论

工业互联网具有区别于传统互联网的多重特点,针对传统入侵检测方法准确性低、难以适应工业互联网组网模式和海量不平衡数据等问题,本文提出一种基于胶囊网络的工业互联网入侵检测方法. 该方法首先参考DRN结构,引入残差块为主胶囊层提取高质量的特征图,然后使用动态路由算法对特征进行聚类,在反向传播中使用Adam算法优化学习率,使检测模型平稳快速收敛,并在气体管道数据集仿真测试中取得99.28%的检测准确率. 即使在数据分布严重不平衡的情况下,实验测试数据结果表明该模型的漏报率、误报率和F1值仍然可以达到1.08%,0.51%和99.03%,较其他对比算法有较大优势,能较好适应工业互联网应用环境.

本文提出的入侵检测方法是针对工业互联网特点构建的,模型基于监督学习模式,需要清晰的数据包信息和流量模式. 虽然可取得较好效果,但传统互联网并不能完全满足这些条件. 同时,动态路由算法复杂度较高,相对于传统方法计算开销较大. 下一步拟对动态路由算法策略进行改进,以减少动态路由消耗时间.

参考文献

[1] DALENOGARE L S, BENITEZ G B, AYALA N F, et al. The expected contribution of Industry 4.0 technologies for industrial performance[J]. International Journal of Production Economics, 2018, 204: 383-394.

[2] ALLADI T, CHAMOLA V, ZEADALLY S. Industrial control systems: Cyberattack trends and countermeasures [J]. Computer Communications, 2020, 155: 1-8.

[3] MARKOVIC-PETROVIC J D, STOJANOVIC M D, RAKAS S V B. A fuzzy AHP approach for security risk assessment in SCADA networks[J]. Advances in Electrical and Computer Engineering, 2019, 19(3): 69-74.

[4] ZOLANVARI M, TEIXEIRA M A, GUPTA L, et al. Machine learning-based network vulnerability analysis of industrial Internet of Things[J]. IEEE Internet of Things Journal, 2019, 6(4): 6822-6834.

[5] TONG D, QU Y R, PRASANNA V K. Accelerating decision tree based traffic classification on FPGA and multi-core platforms[J]. IEEE Transactions on Parallel and Distributed Systems, 2017, 28(11): 3046-3059.

[6] ALVES T, DAS R, MORRIS T. Embedding encryption and machine learning intrusion prevention systems on programmable logic controllers[J]. IEEE Embedded Systems Letters, 2018, 10(3): 99-102.

[7] 江泽涛, 周谭盛子, 韩立尧. 基于感知哈希矩阵的最近邻入侵检测算法[J]. 电子学报, 2019, 47(7): 1538-1546.
JIANG Z T, ZHOU T S Z, HAN L Y. Nearest neighbor intrusion detection method based on perceived hash matrix [J]. Acta Electronica Sinica, 2019, 47(7): 1538-1546. (in Chinese)

[8] POTLURI S, HENRY N F, DIEDRICH C. Evaluation of

hybrid deep learning techniques for ensuring security in networked control systems[C]//2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation(ETFA). Limassol: IEEE, 2017: 1-8.

[9] 王勇, 周慧怡, 俸皓, 等. 基于深度卷积神经网络的网络流量分类方法[J]. 通信学报, 2018, 39(1): 14-23.
WANG Y, ZHOU H Y, FENG H, et al. Network traffic classification method basing on CNN[J]. Journal on Communications, 2018, 39(1): 14-23. (in Chinese)

[10] 陈红松, 陈京九. 基于循环神经网络的无线网络入侵检测分类模型构建与优化研究[J]. 电子与信息学报, 2019, 41(6): 1427-1433.
CHEN H S, CHEN J J. Recurrent neural networks based wireless network intrusion detection and classification model construction and optimization [J]. Journal of Electronics and Information Technology, 2019, 41(6): 1427-1433. (in Chinese)

[11] HE Y B, MENDIS G J, WEI J. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism[J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2505-2516.

[12] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Neural Information Processing Systems. California: NIPS Proceeding, 2017: 3856-3866.

[13] DENG F, PU S L, CHEN X H, et al. Hyperspectral image classification with capsule network using limited training samples[J]. Sensors, 2018, 18(9): 3153.

[14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770-778.

[15] 赵耀霞, 吴桐, 韩焱. 基于卷积神经网络的复杂构件内部零件装配正确性识别[J]. 电子学报, 2018, 46(8): 1983-1988.
ZHAO Y X, WU T, HAN Y. Nearest identifying the correctness of fit of internal components based on a convolutional neural network [J]. Acta Electronica Sinica, 2018, 46(8): 1983-1988. (in Chinese)

[16] DONG X Y, HUANG J S, YANG Y, et al. More is less: A more complicated network with less inference complexity[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2017: 5840-5848.

[17] LI J Q, YU F R, DENG G, et al. Industrial Internet: A survey on the enabling technologies, applications, and challenges[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1504-1526.

[18] 魏琴芳, 吕博文, 胡向东. 基于稀疏化LSSVM的物联网轻量级入侵检测方法[J]. 重庆邮电大学学报(自然科学版), 2021, 33(3): 475-481.
WEI Q F, LV B W, HU X D. A lightweight intrusion detection method for the Internet of things based on sparse LSSVM[J]. Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition), 2021, 33(3): 475-481. (in Chinese)

[19] MANTERE M, SAILIO M, NOPONEN S. Network traffic features for anomaly detection in specific industrial control system network[J]. Future Internet, 2013, 5(4): 460-473.

[20] MORRIS T, GAO W. Industrial control system traffic data sets for intrusion detection research[C]//International Conference on Critical Infrastructure Protection. Berlin: Springer, 2014: 65-78.

[21] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Ottawa: IEEE, 2009: 1-6.

[22] KRAWCZYK B. Learning from imbalanced data: Open challenges and future directions[J]. Progress in Artificial Intelligence, 2016, 5(4): 221-232.

作者简介

胡向东　男,1971年生,四川武胜人. 博士,重庆邮电大学教授,博士生导师. 主要研究方向为智能感知、网络化测量及工业互联网安全等.
E-mail: huxd@cqupt.edu.cn

李之涵　男,1993年生,安徽淮北人. 硕士研究生. 主要研究方向为工业互联网安全.
E-mail: lizhihan_cn@foxmail.com