计算机科学
COMPUTER SCIENCE

一种基于GRU的半监督网络流量异常检测方法

李海涛, 王瑞敏, 董卫宇, 蒋烈辉

引用本文

李海涛, 王瑞敏, 董卫宇, 蒋烈辉.一种基于GRU的半监督网络流量异常检测方法[J]. 计算机科学, 2023, 50(3): 380-390.

LI Haitao, WANG Ruimin, DONG Weiyu, JIANG Liehui. Semi-supervised Network Traffic Anomaly Detection Method Based on GRU [J]. Computer Science, 2023, 50(3): 380-390.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220100032

一种基于GRU的半监督网络流量异常检测方法

李海涛[1] 王瑞敏[2] 董卫宇[2] 蒋烈辉[2]
1 郑州大学网络空间安全学院 郑州 450001
2 信息工程大学数学工程与先进计算国家重点实验室 郑州 450001
(926206615@qq.com)

摘 要 入侵检测系统(IDS)是在出现网络攻击时能够发出警报的检测系统,检测网络中未知的攻击是IDS面临的挑战。深度学习技术在网络流量异常检测方面发挥着重要的作用,但现有的方法大多具有较高的误报率且模型的训练大多使用有监督学习的方式。为此,提出了一种基于门循环单元网络(GRU)的半监督网络流量异常检测方法(SEMI-GRU)。该方法将多层双向门循环单元神经网络(MLB-GRU)和改进的前馈神经网络(FNN)相结合,采用数据过采样技术和半监督学习训练方式,应用二分类和多分类方式检验网络流量异常检测的效果,并使用NSL-KDD,UNSW-NB15和CIC-Bell-DNS-EXF-2021数据集进行验证。与经典机器学习模型和DNN,ANN等深度学习模型相比,SEMI-GRU方法在准确率、精确率、召回率、误报率和F1分数等指标上的表现均表现更优。在NSL-KDD二分类和多分类任务中,SEMI-GRU在F1分数指标上领先于其他方法,分别为93.08%和82.15%;在UNSW-NB15二分类和多分类任务中,SEMI-GRU在F1分数上的表现优于对比方法,分别为88.13%和75.24%;在CIC-Bell-DNS-EXF-2021轻文件攻击数据集二分类任务中,SEMI-GRU对所有测试数据均分类正确。

关键词:入侵检测系统;半监督学习;多层双向门循环单元;前馈神经网络;NSL-KDD;UNSW-NB15
中图法分类号 TP181

Semi-supervised Network Traffic Anomaly Detection Method Based on GRU

LI Haitao[1],WANG Ruimin[2],DONG Weiyu[2] and JIANG Liehui[2]
1 School of Cyber Science and Engineering,Zhengzhou University,Zhenzhou 450001,China
2 State Key Laboratory of Mathematical Engineering and Advanced Computing,Information Engineering University,Zhenzhou 450001,China

Abstract Intrusion detection system(IDS) is a detection system that can issue an alarm when a network attack occurs. Detecting unknown attacks in the network is a challenge that IDS faces. Deep learning technology plays an important role in network traffic anomaly detection,but most of the existing methods have a high false positive rate and most of the models are trained using supervised learning methods. A gated recurrent unit network(GRU)-based semi-supervised network traffic anomaly detection method(SEMI-GRU) is proposed,which combines a multi-layer bidirectional gated recurrent unit neural network(MLB-GRU) and an improved feedforward neural network(FNN). Data oversampling technology and semi-supervised learning training method are used to test the effect of network traffic anomaly detection using binary classification and multi-classification methods,and NSL-KDD,UNSW-NB15 and CIC-Bell-DNS-EXF-2021 datasets are used for verification. Compared with classic machine learning models and deep learning models such as DNN and ANN,the SEMI-GRU method outperforms the machines lear-ning and deep learning methods listed in this paper in terms of accuracy,precision,recall,false positives,and F1 scores. In the NSL-KDD binary and multi-class tasks,SEMI-GRU outperforms other methods on the F1 score metric,which is 93.08% and 82.15%,respectively. In the UNSW-NB15 binary and multi-class tasks,SEMI-GRU outperforms the other methods on the F1 score,which is 88.13% and 75.24%,respectively. In the CIC-Bell-DNS-EXF-2021 light file attack dataset binary classification task,all test data are classified correctly.

Keywords Intrusion detection system,Semi-supervised learning,Multilayer bidirectional GRU,Feedforward neural network,NSL-KDD,UNSW-NB15

1 引言

随着物联网设备和网络用户的持续增长,新型通信技术也在不断发展,计算机网络的规模变得更加庞大和复杂[1-2],网络流量异常检测技术也面临着更大的挑战。

近年来,入侵检测系统(Intrusion Detection System,IDS)

到稿日期:2022-01-04 返修日期:2022-07-17
基金项目:国家重点研发计划(2018YFB0804500)
This work was supported by the National Key R&D Program of China(2018YFB0804500).
通信作者:王瑞敏(380430313@qq.com)

在网络流量异常检测方面发挥了重要作用,其主要用于对网络环境或者主机进行监控并分析,找出其中存在的威胁系统安全的不友好行为。

根据数据来源可以将 IDS 分为基于主机的入侵检测系统(Host based Intrusion Detection System,HIDS)和基于网络的入侵检测系统(Network based Intrusion Detection System,NIDS)。HIDS 可以处理主机操作系统中产生的各种日志信息和审计信息[3]。NIDS 部署在包含多个主机的网络环境中,用于从网络中捕获的数据包中提取信息,或者从网络数据流中提取信息。在学术界,NIDS 被划分为基于误用的 NIDS(MNIDS)和基于异常的 NIDS(ANIDS)[4]。

根据检测方法可将 IDS 分为基于签名的入侵检测系统(Signature based Intrusion Detection System,SIDS)和基于异常的入侵检测系统(Anomaly based Intrusion Detection System,AIDS)。SIDS 采用模式匹配技术发现已知的攻击[5],但需要提前将攻击的规则维护到系统的数据库中。例如 Snort 的运行就需要用到提前定义好的规则文件,根据规则进行攻击发现。尽管 SIDS 检测具有较高的准确度,但无法检测未知的攻击,因此无法完成零日(0-day)漏洞攻击的检测。

近年来 0-day 漏洞增长速率越来越高,AIDS 作为一种可行的解决方案能把偏离正常模式的行为判定为异常或者入侵,在训练阶段对正常行为模式进行建模,在测试阶段将偏离正常行为模式的网络流量数据检测为异常流量。AIDS 可使用基于统计的方法和基于机器学习的方法来建立正常行为的模式[6],机器学习方法的优点是可以建立一种映射关系,更易发现一些隐藏的网络异常,可分为监督和无监督两种学习方式。有监督方式处理有标记的训练集,无监督方式(如聚类分析和主成分分析等)处理无标记的训练集。采用传统的机器学习算法进行异常流量检测时会面临一些困难,如较高的误报率、特征选择困难以及不能有效表示数据之间复杂的非线性关系等。

如今,深度学习方法被广泛应用于医疗、目标检测[7]和自然语言处理[8]等领域,卷积神经网络和循环神经网络已经被应用在异常流量检测方面,但是存在诸多问题。在使用循环神经网络(RNN)的方案中,大多研究[9-10]使用的长短期记忆(LSTM)网络比门循环单元(Gated Recurrent Unit,GRU)网络的参数数量更多,计算时间更长,复杂度更高。许多异常检测数据集存在类别不平衡现象,且各个类别样本数量存在显著差异,这可能会导致无法有效地从少数类样本中提取有效特征。很多研究使用有监督学习的方式,但这种方式依赖大量标记样本,当某些类别标记样本较少时会给模型的训练带来挑战。采用半监督学习方式训练既使用带标签的数据,也使用无标签的数据,同样可以达到有监督学习方法的效果,但是对这种方式的研究还很少。LSTM 是根据历史输入和当前输入来预测未来的特性,由于其具有记忆长时依赖的特性,因此被广泛用于在异常流量检测领域。GRU 作为 LSTM 的简化版本,参数数量更少,同样可达到和 LSTM 相似的效果,在异常流量检测领域具有一定的研究空间。

目前网络流量异常检测的准确率还有提升的空间,误报率也未降低到一个可接受的范围,准确识别一个未知的攻击还存在一定的困难。本研究的目的是使用改进的 GRU 神经网络[11]来进行异常流量检测,提升对异常流量检测的准确率,降低误报率,识别未知攻击。本文提出了一种基于门循环单元神经网络(GRU)的半监督网络流量异常检测方法(SEMI-GRU),将简单的 GRU 神经网络设置为多层(Multi-layer)和双向(Bidirectional),形成了多层双向 GRU 神经网络(MLB-GRU),再利用前馈神经网络(Feed-forward Neural Network,FNN)和 MixMatch[12]对 MLB-GRU 进行优化。本文提出的方法首先使用 SMOTE(Synthetic Minority Over-sampling Technique)算法[13]对训练样本进行数据增强,然后采用改进的前馈神经网络调整 MLB-GRU 输出向量的维度,采用改进的超强半监督学习模型 MixMatch 在总体的损失函数中引入了半监督损失项。最后用新南威尔士大学 2015 网络(UNSW-NB15)数据集[14]、数据库网络安全与实验室知识发现(NSL-KDD)数据集[15]和 CIC-Bell-DNS-EXF-2021 数据集[16]检验所提方法的效果。

2 相关工作

IDS 可以从网络流量中发现网络中的漏洞攻击、DDoS 攻击和病毒传播等。异常流量检测在 IDS 中发挥着巨大的作用。近年来对异常流量检测技术的研究主要使用机器学习的技术,深度学习作为机器学习的重要组成部分可以被用于入侵检测系统,从而抵御网络攻击。下文将列举与本文方法相关的用于异常流量检测的机器学习方法。

传统的异常检测方法包括单类 SVM[17]、最近邻[18]和聚类[19],但是这些方法不适用于高维的流量数据。对于异常检测,无监督的基于深度学习的方法包括深度置信网络[20]、变分自动编码器[21]和对抗自动编码器[22]。但是,为复杂的高维数据开发有效的异常检测方法仍然是一项挑战[23]。

Radford 等[24]利用循环神经网络训练得到的模型来表示网络上计算机之间的通信序列,并用于识别异常网络流量。他们使用 LSTM 来对流序列建模,但是仅仅使用流序列中的源/目的端口号进行序列建模。该方法只能根据端口序列中的前 10 个端口号来预测下一个端口号,应用的范围有限。Wang 等[25]提出使用深度神经网络来学习分层时空特征,从而改善入侵检测。他们通过改进文献[26]中的研究提出了分层时空特征入侵检测系统(HAST-IDS),将处理的粒度从 IP 流转到 IP 包,但是训练数据的不平衡导致 HAST-IDS 无法学习到足够的代表性特征。

Vinayakumar 等[27]提出了一种混合的深度神经网络(DNN)结构用于入侵检测,其能够对大规模的数据量进行实时分析。该方法对 NSL-KDD 数据集多分类和二分类测试的准确率分别为 78.50%和 80.10%,对 UNSW-NB15 测试的准确率分别为 66.00%和 78.40%。

Javaid 等[28]提出了一种自学习(STL)深度学习模型用于入侵检测。该模型通过自动编码器(AE)进行无监督地预处理,通过人工神经网络(ANN)和 Softmax 实现分类;使用 NSL-KDD 数据集进行多分类和二分类的异常检测,测试的准确率分别为 79.10%和 88.39%。Ingre 等[29]使用 ANN 网络结构设计了一种入侵检测系统,在 NSL-KDD 数据集上

多分类和二分类测试的准确率分别为 79.90%和 81.20%。

一些研究关注到了类别不平衡问题，Wu 等[30]构造了 CNN 和 RNN 进行 NSL-KDD 数据集分类的研究。他们提出了一种代价函数来解决样本类别不平衡的问题，代价函数的权重系统根据训练集中各个类别样本的数量来调整。在他们的实验中，CNN 的效果要略差于 RNN，对 NSL-KDD 数据集多分类测试的准确率为 79.48%。Al-Turaiki 等[31]提出了一种 CNN 网络结构，使用 SMOTE 进行样本过采样，使用 Deep Feature Synthesis(DFS)构建特征工程来从 NSL-KDD 数据集中获取更多的特征信息；创新性地提出了一种跳跃连接(Skip Connection)用于连接 CNN 中不同的卷积层，在测试集上进行多分类和二分类地异常流量检测，取得了不错的效果。对 NSL-KDD 数据集多分类和二分类测试的准确率分别为 81.10%和 88.81%，对 UNSW-NB15 数据集多分类和二分类测试的准确率分别为 69.46%和 90.25%。Altwaijry 等[32]提出了使用 DNN 网络结构进行 NSL-KDD 数据集和 UNSW-NB15 数据集二分类和多分类的异常流量检测，对 NSL-KDD 数据集多分类和二分类测试的准确率分别为 77.55%和 84.70%，对 UNSW-NB15 数据集多分类和二分类测试的准确率分别为 62.87%和 80.63%。Xu 等[33]通过改进的自动编码器网络对 NSL-KDD 数据集进行入侵检测研究，对该数据集二分类测试的准确率为 90.61%。Raj 等[34]使用分类提升技术来进行 NSL-KDD 数据集二分类流量检测，识别的准确率为 99.92%，但是精确率和召回率分别为 79.42%和 80.00%。

3 研究方法

RNN 的关键优势是具有记忆特性，可以根据上一时间步的输出向量和当前时间步的输入向量来决定当前时间步的输出，模型的参数会随着每个时间步的输入而更新。普通 RNN 的缺点是不能记忆长时依赖，固有的遗忘性导致其只能够记录短时依赖，这会给模型的训练带来负面的影响。RNN 中有两种不同的实现方式，分别是 LSTM 和 GRU。LSTM 作为 RNN 的一种实现方式，其内部具有输入门、遗忘门和输出门 3 个控制单元。输入门控制着网络的输入，遗忘门控制着记忆单元，输出门控制着网络的输出。正是由于遗忘门的作用，使得 LSTM 具有了长时记忆的功能。GRU 和 LSTM 最大的不同在于 GRU 将遗忘门和输入门合并成了一个“更新门”，网络不再给出记忆状态，而是将输出结果作为记忆状态不断向后循环传递，使网络的输入和输出都变得更简单。

3.1 SEMI-GRU 方法总体架构

本文在 GRU 模型的基础上，将 GRU 的层数和方向看作可以调整的参数，因此形成的 MLB-GRU 模型比 GRU 模型的表达能力更强并在此基础上提出了一种数据预处理方法，将输入数据转为二进制特征，接着利用 SMOTE 方法进行少数类样本过采样，然后使用 MLB-GRU 和提出的对称缩减 FNN 网络结构(Symmetric Reduction FNN，SR-FNN)进行特征提取，最后使用所提出的 MixMatch 方案简化版向总体损失函数中引入半监督损失项，最终形成了 SEMI-GRU 方法。所提出的 SEMI-GRU 方法由 3 个主要部分组成，分别是 MLB-GRU，SR-FNN 以及 MixMatch 方案简化版，如图 1 所示。

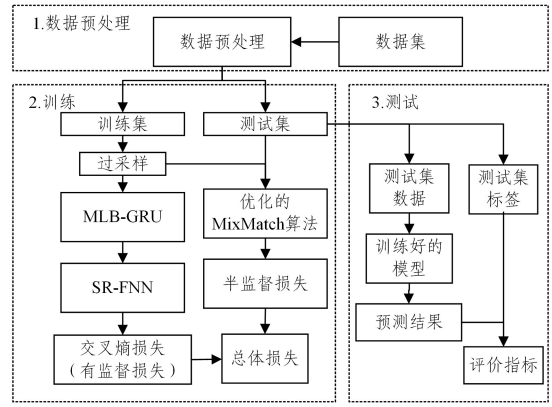

图 1 SEMI_GRU 方法总体架构

Fig. 1 Overall architecture of SEMI_GRU method

SEMI-GRU 包含 3 个主要的阶段，分别是数据预处理、训练和测试阶段。在训练阶段进行数据预处理，预处理完成之后每条样本将占用 361 字节的内存大小。为了将预处理好的数据输入 GRU 神经网络中，我们把每条数据对应的输入向量划分为 19 个时间步，每个时间步的输入大小为 19。在训练阶段，将经过 SMOTE 过采样的训练集输入模型中可以得到交叉熵损失，同样，将经过 SMOTE 过采样的训练集和测试集输入简化的 MixMatch 简化版模型中可以得到半监督损失。我们将交叉熵损失和半监督损失作为总的损失函数，使模型朝着最小化损失函数的方向进行训练。在测试阶段，将测试集输入训练好的模型中，得到预测的标签。将预测的标签和真实的标签进行对比，可以得到预测的结果。

3.2 数据集描述

本研究采用两种经典数据集和一种较新的现实世界网络中的数据集，分别是 NSL-KDD 数据集[15]、UNSW-NB15 数据集[14]和 CIC-Bell-DNS-EXF-2021 数据集[16]。

3.2.1 NSL-KDD 数据集

NSL-KDD 数据集是用于评估入侵检测系统的一个基准数据集，是 KDD99 数据集的改进版，它弥补了 KDD99 数据集的天然缺陷。NSL-KDD 数据集的训练集中不包含冗余记录[15]，测试集中没有重复的记录，训练集和测试集样本数量设置合理。相比 KDD99 数据集，NSL-KDD 数据集更能衡量一个深度学习模型的优劣，表 1 列出了 NSL-KDD 攻击类型。

表 1 不同攻击类别下攻击类型

Table 1 Attack types in different attack categories

类别	训练集攻击类型	测试集特有的攻击类型
Dos	back, neptune, smurf, teardrop, land, pod	apache2, mailbomb, processtable
Probe	satan, portsweep, ipsweep, nmap	mscan, saint
R2L	warezmaster, warezclient, ftp-write, guesspassword, imap multihop, phf, spy	sendmail, named, worm, snmpgetattack, xlock, xsnoop snmpguess
U2R	rootkit, bufferoverflow, load-module, perl	httptunnel, ps, sqlattack, xterm

该数据集包含四大攻击类别，分别是拒绝服务(DoS)、探针攻击、用户到 root 攻击(U2R)和远程到本地攻击(R2L)。训练集有 22 种攻击类型，测试集有 38 种攻击类型(包括测试集特有的 16 种攻击类型)。该数据集中共包含 41 个特征，

其中有 3 个离散特征：协议类型（protocol_type）、服务（service）和标记（flag）字段。41 个特征可以分为 3 组：根据传输层协议提取的基本特征字段、根据窗口间隔提取的流量特征字段和根据应用层连接数据提取的内容特征字段。如表 2 所列，在训练集中包含 125 973 条记录，在测试集中包含 22 544 条记录。

表 2 各攻击类别样本数量统计

Table 2 Statistics of the number of samples in each attack type

数据集	Normal	Dos	Probe	R2L	U2R	Total
Train	67 343	45 927	11 656	995	52	125 973
Test	9 711	7 458	2 421	2 754	200	22 544

3.2.2 UNSW-NB15 数据集

尽管 NSL-KDD 数据集可以作为入侵检测系统的评估数据集，但是该数据集被认为是一种过时的数据集。UNSW-NB15 数据集[14]由澳大利亚网络安全中心（ACCS）于 2015 年创建，其中包含最新的攻击类型。其创建过程为：首先使用 Tcpdump 工具捕获由现实世界正常网络活动和人工发起的短暂攻击行为所产生的流量数据；然后使用 Bro-IDS 工具和算法处理捕获的流量数据，生成含有 48 个特征的 CSV 文本文件；最后从文本文件中抽取一部分数据得到含有 44 个特征、175 341 条记录的训练集（UNSW_NB15_training-set. csv）和含有 44 个特征、82 332 条记录的测试集（UNSW_NB15_testing-set. csv）。该数据集共有 9 种不同的攻击类型，分别是：Fuzzers，Analysis，Backdoors，DoS，Exploits，Generic，Reconnaissance，Shellcode 和 Worms。UNSW-NB15 训练集的统计信息如表 3 所列，测试集的统计信息如表 4 所列。

表 3 UNSW-NB15 训练集各类别统计信息

Table 3 Statistics of UNSW-NB15 training set for each category

类型	数量	类型	数量
Benign	56 000	Reconnaissance	10 491
Generic	40 000	Analysis	2 000
Exploits	33 393	Backdoor	1 746
Fuzzers	18 184	Shellcode	1 133
DoS	12 264	Worms	130

表 4 UNSW-NB15 测试集各类别统计信息

Table 4 Statistics of UNSW-NB15 test set for each category

类型	数量	类型	数量
Normal	37 000	Reconnaissance	3 496
Generic	18 871	Analysis	677
Exploits	11 132	Backdoor	583
Fuzzers	6 062	Shellcode	378
DoS	4 089	Worms	44

3.2.3 CIC-Bell-DNS-EXF-2021 数据集

CIC-Bell_NDS-EXF-2021 数据集是一个包含 270.8 MB DNS 流量的大型数据集，是通过泄露不同大小和类型的文件产生的。该数据集由 DNS 数据泄露攻击流量构成，包括轻文件攻击和重文件攻击两类攻击。轻文件攻击和重文件攻击均有 6 种文件类型，包括音频、压缩文件、. exe、图形、文本和视频。不失一般性，本文采用轻文件攻击过程产生的数据包构成的 pcap 文件作为源数据来构造本文用到的数据集。轻文件攻击过程产生了 7 个 pcap 文件，分别是 benign. pcap，light_audio. pcap，light_compressed. pcap，light_exe. pcap，light_image. pcap，light_text. pcap 和 light_video. pcap。我们使用工具从 pcap 文件中按照五元组提取出所有的单向流，构造出训练集和测试集，设置训练集和测试集样本数目比例为 9∶1，轻文件攻击数据集中各攻击类别样本数统计如表 5 所列。

表 5 轻文件攻击数据集各攻击类别样本数统计

Table 5 Statistics of the number of samples of each attack category in light file attack data set

类别	训练集	测试集	总体数量
Benign	31 667	3 519	35 186
Light_exe	14 804	1 645	16 449
Light_compressed	20 965	2 330	23 295
Light_video	11 079	1 231	12 310
Light_text	7 171	797	7 968
Light_audio	17 789	1 976	19 765
Light_image	1 359	151	1 510

3.3 数据预处理

数据预处理是将原始的数据集转换为被神经网络处理的数值类型的数据。对于 RNN 网络，输入数据的维度是[*num_steps*，*input_dim*]，其中 *num_steps* 表示循环输入的次数，*input_dim* 表示输入向量的维度。NSL-KDD 数据集含有 3 个离散的特征：protocol_type，service 和 flag。UNSW-NB15 数据集同样存在 3 个离散的特征：proto，service 和 state。对于离散的特征，可将离散的值映射为连续的自然数，互不相同的离散值的个数等于自然数的个数。由于自然数是整型的数值，因此可输入到神经网络中。

NSL-KDD 数据集含有 42 个特征，UNSW-NB15 数据集含有 43 个特征（将 label 字段和 atttack_cat 字段合并之后）。将离散的特征映射为连续的自然数之后，数据集中所有字段被转换为数值类型：Int64 和 Float64，每个数值占用 8 个字节。因此，NSL-KDD 数据集中每个样本的 42 个字段会占用 336（42 * 8）个字节内存，UNSW-NB15 数据集中每个样本的 43 个字段会占用 344（43 * 8）字节内存。由于 336 和 344 均小于 361（19 * 19），因此可将 NSL-KDD 和 UNSW-NB15 数据集中任意一个样本映射为一个 19 * 19 的单通道二维像素图，其中每个像素占用一个字节。像素图可以被存储起来作为神经网路的输入，在加载数据集时可以将其转换为输入向量。对于像素图末尾的空缺字节，可以使用数值 128 进行填充。由于一个字节的数值范围是 0～255，因此使用 128 填充空缺的字节将会获得更大的容错率。对于 NSL-KDD 数据集，需填充 25 字节的内容；对于 UNSW-NB15 数据集，需填充 17 字节的内容。

3.4 样本过采样

当训练集存在类别不平衡现象时，模型不能充分地从含有少量样本的类别中学习到该类别的特征。深度神经网络更倾向于从含有较多样本的类别中学习，样本过采样就是在不影响其他类别的情况下，增加少数类的样本量。通过样本生成的方式进行过采样，包括使用无监督的自动编码器网络、生成对抗网络或者翻转变换等。但是这些方法的效果并不理想。随机过采样的方式是先从少数类的样本中随机抽样，再将抽样样本添加到数据集中。这种方法的优点是简单，缺点

是容易造成过拟合。合成少数类过采样技术(SMOTE)是基于随机过采样的一种改进方案,其基本思想是对少数类样本进行插值来人工合成新样本,并将其添加到数据集中。算法流程如下:

(1)对于少数类中的每一个样本,以欧氏距离计算它到少数类样本集中所有样本的距离,得到其 k 近邻。

(2)根据样本不平衡比例设置一个采样比例以确定采样倍率 N,对于每一个少数类样本 x_n,从其 k 近邻中选择若干个样本,假设选择的近邻为 o。

(3)对于每个随机选出的近邻 o,分别与原样本按照式(1)计算得到新产生的样本。

$$x_{new}=o+rand(0,1)*|o-x_n| \tag{1}$$

如表 2 所列,对 NSL-KDD 数据集的训练集进行分析,R2L 攻击类别包含 995 个样本,U2R 攻击类别包含 52 个样本,相比其他攻击类别而言攻击样本较少。本文使用 SMOTE 算法分别对这两种攻击类别进行处理,对 R2L 攻击样本过采样为原来的 10 倍,对 U2R 攻击样本过采样为原来的 100 倍。处理完成之后,R2L 攻击类别含有 10000 个样本,U2R 攻击类别含有 5 200 个样本。对于 UNSW-NB15 数据集,不对其进行过采样处理。通过实验我们发现对 UNSW-NB15 数据集采用过采样处理并不会显著提升模型的效果,这可能是因为该数据集的训练集和测试集数据分布相似性不高。

3.5 多层双向 GRU 模型(MLB-GRU)

GRU 相比 LSTM 的优势在于它能够以更少的参数量实现捕获长时依赖的功能,模型的训练速度更快。

图 2 为多层 GRU 神经网络示意图,这里输入的时间步是从左向右进行的。图中展示的是单向的输入,实际上输入可以是双向的,即输入向量按照从右向左的顺序输入。不失一般性,图中展示的 GRU 的层数为 2,在实际训练的过程中可以对层数进行调整。双向的 GRU 模型,每个方向都会得到一组输出,这两组输出在 GRU 模型的最后一层可以合并在一起构成多层双向 GRU 模型的输出向量。

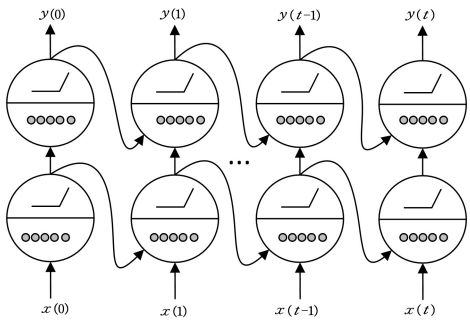

图 2 多层 GRU 神经网络

Fig. 2 Multi-layer GRU neural network

假设前一时间步是 $t-1$,当前时间步是 t,当前时间步的输入是 $x_{(t)}$,上一时间步的输出是 $y_{(t-1)}$,那么当前时间步的输出 $y_{(t)}$ 的计算式如式(2)所示。将遗忘门和输入门进行合并,得到更新门,再根据更新门和上一时间步的输出得到当前时间步的输出。

$$\begin{cases} z_t=\sigma(W_z\cdot[y_{t-1},x_t]) \\ r_t=\sigma(W_r\cdot[y_{t-1},x_t]) \\ \tilde{y}_t=\sigma(W\cdot[r_t*y_{t-1},x_t]) \\ y_t=(1-z_t)*y_{t-1}+z_t*\tilde{y}_t \end{cases} \tag{2}$$

将所有时间步的 y_t 拼接在一起,就构成了输出向量。由于是双向的 GRU 神经网络,假设前向的输出向量为 $\overrightarrow{output}$,反向的输出向量为 $\overleftarrow{output}$,那么最终的输出向量如式(3)所示:

$$output=\text{concat}(\overrightarrow{output},\overleftarrow{output}) \tag{3}$$

3.6 对称缩减 FNN 网络结构

本文设计了一种前馈神经网络结构,根据其结构特征将其命名为对称缩减 FNN 网络(Symmetrical Reductional feed Forward Neural Network,SR-FNN)。在该网络内部,参数从输入层经过隐含层向输出层进行单向传播。与循环神经网络不同,其内部不会构成有向环,可将上层 GRU 网络模型的输出向量输入到该结构中,从而调整输出向量的维度。SR-FNN 的网络结构如图 3 所示。

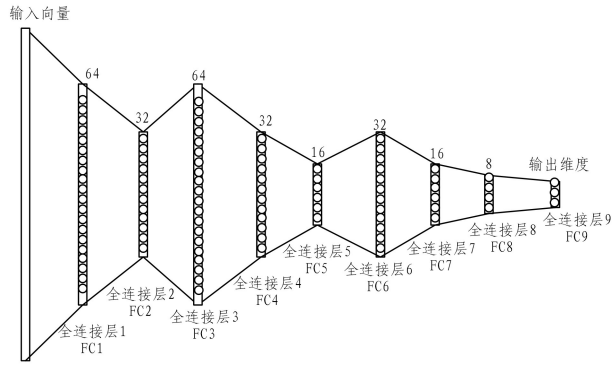

图 3 SR-FNN 网络结构

Fig. 3 SR-FNN network structure

从图 3 可以看出,所提出的 FNN 网络结构总共包含 9 个全连接层,其中输入向量来自上层 GRU 神经网络的输出。该 FNN 结构包含 3 组分别由 3 个全连接层拼接在一起的网络结构,第一组网络结构包含 FC1,FC2 和 FC3,第二组包含 FC4,FC5 和 FC6,第三组包含 FC7,FC8 和 FC9。其中,第一组网络结构中 FC1 和 FC3 的输出维度相同,第二组中 FC4 和 FC6 的输出维度相同,第三组中 FC7,FC8 和 FC9 的输出维度是递减的。将第三组网络结构中最后一个全连接层 FC9 的输出维度设置为网络流量类别的个数,可以观察到,第一组和第二组网络结构的输出维度都是先减小为原来的一半,然后再恢复到以前的大小,实验表明这种维度先减小后恢复的结构确实提高了模型的效果。

3.7 半监督训练

半监督学习是人工智能发展的一种趋势,半监督学习可以用大量未标记数据和极少量的标记数据学习。Berthelot 等[12]提出的 MixMatch 方法仅使用少量标记数据,就使半监督学习的预测精度逼近有监督学习。MixMatch 方法把多种半监督方法的优势集成在一起,如支持自洽正则化、最小化未标记数据的熵、使用 Weight decay 代替 L2 正则化,以及使用

了 Mixup[35]这种数据增广方法。

基于 MixMatch 方法,我们提出了一种 MixMatch 方法的简化版本,从而在总体损失函数中添加一个半监督损失项,提升模型的泛化能力和效果。在原 MixMatch 方案中,对一个训练批次(Batch)的标记数据 x 和一个 Batch 的未标记数据 u 进行增广,分别得到 1 个 Batch 的增广数据 x' 和 K 个 Batch 的 u'。在本文中,在进行数据增广之前将标记数据的 Batch 和未标记数据的 Batch 设置为相等,从而在未标记数据增广操作完成之后,无须再将其平均划分为 K 个部分,具体过程如算法 1 所示。

算法 1 优化的 MixMatch 算法

Input:x,y,u

Output:loss

1. for(x_b,y_b)in(x,y)do
2. p_b=One-hot(y_b)
3. u_b=next(u)
4. $\bar{q}_b$=model(y|u_b;θ)
5. $q_b=\frac{\text{softmax}(\bar{q}_b)^2}{\text{sum}(\text{softmax}(\bar{q}_b)^2)}$
6. inputs=concatenate(x_b,u_b)
7. m_inputs=MixUp(inputs,shuffle(inputs))
8. targets=concatenate(p_b,q_b)
9. m_targets=MixUp(targets,shuffle(targets))
10. logits_x=model(y|m_inputs[0])
11. logits_u=model(y|m_inputs[1])
12. $loss_1$=−m_targets[0] * log(softmax(logits_x))
13. $loss_2$=MSE(logits_u,m_targets[1])
14. loss=$loss_2$+$loss_2$
15. return loss
16. endfor

从算法 1 可以看出,在所提出的简化版的 MixMatch 方案中,将有标签的训练数据和无标签的测试数据相结合,输入到深度神经网络中,再将输出的内容和原先的标签一起输入到损失函数中。在算法的末尾会将得到的半监督损失项返回,从而使半监督损失和有监督损失一起被优化,这增强了模型的泛化能力和效果。

4 实验结果

4.1 实验设置

实验采用的软件环境是 Ubuntu18.04 操作系统,Pytorch1.9.1 和 Cuda9.0。硬件环境是 Genuine Intel(R) CPU@2.00GHz,32GB 内存,Tesla K80 GPU 11GB 显存。

4.2 超参数设置

(1)学习率的设置。本文根据数据集来设置对应的学习率,这样模型能够取得更好的效果。当使用 NSL-KDD 数据集训练时,学习率设置为 0.002,经过实验验证,该值可以使得训练过程很快收敛且不会陷入局部最优解的状态。当使用 UNSW-NB15 数据集时,我们采用 3 种不同的学习率:0.002,0.02 和 0.01。

(2)批大小的设置。本文在进行 NSL-KDD 数据集多分类时,将批大小设置为 128;在进行 NSL-KDD 数据集二分类异常流量检测时,将批大小设置为 512;当使用 UNSW-NB15 数据集时,将学习率设置为 128。

(3)GRU 神经网络层数的设置。在实验中我们发现,如果 GRU 的层数设置得过大,如设置为 10,模型的训练过程就会很快陷入过拟合状态。如果层数设置得过小,如设置为 1,模型的收敛速度将会很慢。对于 NSL-KDD 数据集,本文将 GRU 的层数设置为 4,5,6,7 和 8 这 5 种不同的值。对于 UNSW-NB15 数据集,本文将 GRU 的层数设置为 3,4,5 和 6 这 4 种不同的值。GRU 有两个方向,在实验中我们发现,双向 GRU 模型的效果优于单向的效果,因此这里将 GRU 网络设置为双向。GRU 的 hidden_size 设置为 64,因此每层的循环单元中神经元的个数为 64。

4.3 评价指标

在深度学习领域,常用的评价指标有准确率、精确率、召回率、F1-Score 和 ROC 曲线,本实验将采用这些指标对 SEMI-GRU 方法进行评价。

TP 指预测为正例且预测正确的数量;TN 指预测为负例且预测正确的数量;FP 指预测为正例且预测错误的数量;FN 指预测为负例且预测错误的数量。

准确率指所有预测正确的正样本、负样本数量之和与总体样本数量的比值,如式(4)所示:

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

对于一个特定的类别,会存在精确率和召回率。精确率也被称为查准率,用于衡量模型找到相关目标的能力,如式(5)所示:

$$Precision=\frac{TP}{TP+FP} \tag{5}$$

召回率也被称为查全率,用于衡量模型找到全部相关目标的能力,即模型给出的预测结果最多能覆盖多少真实目标,如式(6)所示:

$$Recall=\frac{TP}{TP+FN} \tag{6}$$

F1-Score 是基于精确率和召回率的调和平均,如式(7)所示:

$$F1=\frac{2\times P\times R}{P+R}=\frac{2TP}{2TP+FP+FN} \tag{7}$$

假正例率(FPR)也被称为误报率,如式(8)所示。在异常流量检测任务中,常常将攻击类别定义为正例,将正常类别的流量定义为负例,误报指把正常流量识别为攻击流量,而这是 IDS 最不能容忍的。

$$FPR=\frac{FP}{FP+TN} \tag{8}$$

4.4 性能评估

本文针对 NSL-KDD 数据集、UNSW-NB15 数据集和 CIC-Bell-DNS-EXF-2021 数据集进行模型的训练和评估。对于 NSL-KDD 数据集,将进行五分类和二分类研究。五分类将细化为 4 种攻击类别,二分类会对正常流量和异常流量进行区分。对于 UNSW-NB15 和 CIC-Bell-DNS-EXF-2021 数据集,将进行多分类和二分类的异常流量识别。

4.4.1 NSL-KDD 数据集结果

表 6 列出了 SEMI-GRU 方法不同 GRU 层数对 NSL-KDD 数据集的五分类性能比较。本文将 GRU 模型设计为不同的层，并分别进行实验。实验结果表明，5 层和 8 层的 GRU 模型表现较好。其中，使用 5 层的 GRU 模型时，测试集的准确率达到了 83.79%，这个值在对比方法中是最高的。使用 8 层的 GRU 模型可以得到更高的 F 分数(82.15%)。F 分数作为精确率和召回率的调和平均，可以和准确率一起作为模型的综合评价指标。随着 GRU 层数的加深，模型的误报率呈逐渐增长的趋势，最终在第 7 层达到了 3.67%。但又在最深的那一层骤降到了 3.15%，这和 4 层的 GRU 模型的 3.14%误报率相差不大。

表 6 SEMI-GRU 方法不同 GRU 层数对 NSL-KDD 数据集的五分类性能比较

Table 6 Five-category performance comparison of SEMI-GRU method with different GRU layers on NSL-KDD data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	False alarm rate/%	Training data set	Testing data set
4-Layer	82.75	91.62	72.07	80.68	3.14	KDDTrain+	KDDTest+
5-Layer	83.79	91.18	73.99	81.69	3.25	KDDTrain+	KDDTest+
6-Layer	83.73	90.16	73.96	81.26	3.36	KDDTrain+	KDDTest+
7-Layer	83.08	91.36	73.05	81.18	3.67	KDDTrain+	KDDTest+
8-Layer	83.31	93.80	73.08	82.15	3.15	KDDTrain+	KDDTest+

表 7 列出了 SEMI-GRU 方法与其他方法在 NSL-KDD 数据集上的五分类性能比较。从表中可知，除了召回率这一单项指标外，本文提出的方法在所有对比方法中可以达到最高的准确率(83.31%)、最高的精确率(93.80%)和最高的 F 值(82.15%)。这表明 SEMI-GRU 在 NSL-KDD 数据集五分类上具有很好的效果。

表 7 SEMI-GRU 方法与其他方法在 NSL-KDD 数据集上的五分类性能比较

Table 7 Five-category performance comparison of SEMI-GRU and other methods on NSL-KDD data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	Training data set	Testing data set
AlertNet[31]	78.50	81.00	78.50	76.50	KDDTrain+	KDDTest+
DNN[32]	79.10	83.00	68.00	75.76	KDDTrain+	KDDTest+
ANN[33]	79.90	—	—	—	KDDTrain+	KDDTest+
CNN[34]	79.48	—	68.66	—	KDDTrain+	KDDTest+
MCNN[35]	81.10	83.00	81.00	80.00	KDDTrain+	KDDTest+
MCNN-DFS[35]	81.44	81.00	84.00	80.00	KDDTrain+	KDDTest+
MDNN[36]	77.55	81.23	77.55	75.43	KDDTrain+	KDDTest+
Naive Bayes	72.73	76.10	72.70	72.60	KDDTrain+	KDDTest+
J48	74.99	79.60	75.00	71.10	KDDTrain+	KDDTest+
Random Forest	76.45	82.10	76.40	72.50	KDDTrain+	KDDTest+
Bagging	74.83	78.30	74.80	71.60	KDDTrain+	KDDTest+
Adaboost	66.43	—	66.00	—	KDDTrain+	KDDTest+
SEMI-GRU	83.31	93.80	73.08	82.15	KDDTrain+	KDDTest+

表 8 列出了 SEMI-GRU 方法不同 batch 大小和 GRU 层数对 NSL-KDD 数据集二分类的性能比较，实验设计了 5 层和 6 层这两种 GRU 网络，128 和 512 这两种批大小，并对不同的批大小和层数进行组合。实验结果表明，同为 5 层的 GRU 网络，批大小设计为 512 比 128 的效果更好。批大小同为 512 时，5 层和 6 层 GRU 模型的效果体现出了某些规律，如 5 层的 GRU 模型的准确率、召回率和 F 值都比 6 层的 GRU 模型更好，但是其误报率(8.16%)也高于 6 层 GRU 模型的误报率(3.80%)。因此可以看出，当模型的效果较好时，模型的异常流量识别能力就会变强，但同时也更容易把正常的流量识别为异常的流量，这会提升入侵检测系统的误报率。因此在表 8 所列的实验结果中，尽管 5 层的 GRU 模型的准确率、召回率和 F 值略高，但 6 层的 GRU 模型效果可能要优于 5 层的 GRU 模型。

表 8 SEMI-GRU 方法不同 batch 和 GRU 层数对 NSL-KDD 数据集的二分类性能比较

Table 8 Two-category performance comparison of SEMI-GRU method with different batch and GRU layers on NSL-KDD data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	False alarm rate/%	Training data set	Testing data set
5 层，batch128	88.13	97.05	81.63	88.67	3.27	KDDTrain+	KDDTest+
5 层，batch512	92.18	93.74	92.43	93.08	8.16	KDDTrain+	KDDTest+
6 层，batch512	90.89	96.80	86.88	91.57	3.80	KDDTrain+	KDDTest+

表 9 列出了 SEMI-GRU 方法与相关方法对 NSL-KDD 数据集的二分类性能比较。可以看出，除了精确率(93.74%)略低于 J48(97.14%)，召回率(92.48%)略低于 AlertNet(96.90%)外，所提方法在准确率和 F 值上的表现均优于所有对比方法，因此 SEMI-GRU 方法的综合表现优于所有对比方法。SEMI-GRU 在 NSL-KDD 数据集上二分类准确率达到了 92.18%，F 值达到了 93.08%。

表 9 SEMI-GRU 方法与相关方法对 NSL-KDD 数据集的二分类性能比较

Table 9 Two-category performance comparison of SEMI-GRU and related methods on NSL-KDD data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	Training data set	Testing data set
AlertNet[31]	80.10	69.20	96.90	80.70	KDDTrain+	KDDTest+
DNN[32]	88.39	85.44	95.95	90.40	KDDTrain+	KDDTest+
ANN[33]	81.20	—	—	—	KDDTrain+	KDDTest+
BCNN[35]	88.81	89.00	89.00	89.00	KDDTrain+	KDDTest+
BCNN-DFS[35]	90.14	90.00	90.00	90.00	KDDTrain+	KDDTest+
BDNN[36]	84.70	79.45	87.00	83.05	KDDTrain+	KDDTest+
Improved Autoncoder[37]	90.61	86.83	98.43	92.26	KDDTrain+	KDDTest+
CATBoost[38]	99.92	79.24	80.00	79.71	KDDTrain+	KDDTest+
Naive Bayes	76.12	92.38	63.27	75.10	KDDTrain+	KDDTest+
J48	81.53	97.14	69.61	81.10	KDDTrain+	KDDTest+
Random Forest	80.45	97.05	67.72	79.77	KDDTrain+	KDDTest+
Bagging	82.63	91.87	76.23	83.32	KDDTrain+	KDDTest+
Adaboost	78.44	95.28	65.37	77.54	KDDTrain+	KDDTest+
SEMI-GRU	92.18	93.74	92.43	93.08	KDDTrain+	KDDTest+

4.4.2 UNSW-NB15 数据集结果

本文针对 SEMI-GRU 方法设置不同的学习率(lr)和 GRU 层数,并对 UNSW-NB15 数据集进行十分类性能比较,结果如表 10 所列。首先将学习率设置为 0.002,然后将 GRU 模型设置为 3,4,5,6 层进行实验。结果显示,当学习率设置为 0.002 且 GRU 网络层数设置为 4 时,SEMI-GRU 方法可以取得较好的效果,达到了最高的准确率(73.70%)和最高的 *F* 值(68.34%)。因此,可以把层数固定为 4、然后使用不同的学习率进行训练,将学习率增大为 0.01 和 0.02,可以看出学习率设置为 0.01 时可以取得更好的效果,准确率达到了 80.58%,召回率达到了 75.24%,为所有对比方法的最高值。适当增大学习率之后,不仅提升了 SEMI-GRU 方法的效果,而且降低了误报率。这说明选择一个合适的学习率进行学习,可以使得模型取得更好的效果。尤其是当 GRU 模型的层数设置为 4、学习率设置为 0.02 时,可以取得最低的误报率 2.30%,因此所提方法满足入侵检测系统误报率较低的核心要求。

表 10 SEMI-GRU 方法不同学习率和层数对 UNSW-NB15 数据集的十分类性能比较

Table 10 Ten-category performance comparison of SEMI-GRU method with different learning rates and layers on UNSW-NB15 data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	False alarm rate/%	Training data set	Testing data set
3 层,0.002(lr)	72.64	69.92	72.66	67.61	4.21	UN-train	UN-test
4 层,0.002(lr)	73.70	67.84	73.70	68.34	4.89	UN-train	UN-test
5 层,0.002(lr)	73.57	68.51	73.57	67.42	3.31	UN-train	UN-test
6 层,0.002(lr)	70.15	65.75	70.15	64.24	3.23	UN-train	UN-test
4 层,0.02(lr)	79.77	72.81	79.77	74.56	2.30	UN-train	UN-test
4 层,0.01(lr)	80.58	76.77	71.92	75.24	3.08	UN-train	UN-test

表 11 列出了 SEMI-GRU 方法与相关方法对 UNSW-NB15 数据集的十分类性能比较。除了精确率外,SEMI-GRU 方法在准确率、召回率和 *F* 值等指标上的表现均优于对比方法。说明 SEMI-GRU 在进行 UNSW-NB15 数据集上的多分类异常检测时可以取得较好的效果。表 12 列出了 SEMI-GRU 方法与相关方法对 UNSW-NB15 数据集二分类的性能比较,可以看到,本文方法在该项任务中的表现优于或接近对比方法。实际上,SEMI-GRU 在进行二分类异常流量检测时对 Benign 类型的流量检测可以实现较高的精确率(99.54%),但是召回率却较低(57.6%);在进行十分类异常流量检测时对 Bengin 类型的流量具有较高的召回率(96.92%)。为了显著提高 SEMI-GRU 在 UNSW-NB15 数据集上进行二分类异常检测的效果,本文将二分类模型和十分类模型相结合进行二分类异常流量检测。测试的一批数据分别输入到对 Benign 类别识别精确率较高的二分类模型中和对 Benign 类别识别召回率较高的十分类模型中,对应两组输出。把第一组表示预测结果的向量中为 0 的元素的索引记录下来,然后将第二组表示预测结果的向量中对应索引位置处的值设置为 0,之后继续对第二组预测向量进行处理,其他索引处对应元素的值如果不为 0(Benign),那么就将其设置为 1(Anomaly)。对每批测试数据都进行此操作,最终可以实现较好的二分类效果。SEMI-GRU 在该项任务中可以获得最高的精确率(96.29%),且本文方法的准确率(88.11%)和在该项任务中最高的准确率(90.25%)非常接近,本文方法的 *F* 值(88.13%)和在该项任务中最高的 *F* 值(90.45%)也很接近。

表 11 SEMI-GRU 方法与相关方法对 UNSW-NB15 数据集的十分类性能比较

Table 11 Ten-category performance comparison of SEMI-GRU and related methods on UNSW-NB15 data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	Training data set	Testing data set
AlertNet[31]	66.00	62.30	66.00	59.60	UN-train	UN-test
MCNN[35]	69.46	84.00	69.00	74.00	UN-train	UN-test
MCNN-DFS[35]	68.52	83.00	69.00	73.00	UN-train	UN-test
MDNN[36]	62.87	76.00	63.00	64.00	UN-train	UN-test
NaiveBayes	45.22	29.67	38.62	33.56	UN-train	UN-test
J48	51.50	28.18	21.48	24.38	UN-train	UN-test
Random Forest	68.09	62.51	35.15	44.99	UN-train	UN-test
Bagging	51.45	32.85	21.45	25.95	UN-train	UN-test
Adaboost	51.50	28.18	21.48	24.38	UN-train	UN-test
SEMI-GRU	80.58	76.77	71.92	75.24	UN-train	UN-test

表 12 SEMI-GRU 方法与相关方法对 UNSW-NB15 数据集的二分类性能比较

Table 12 Two-category performance comparison of SEMI-GRU and related methods on UNSW-NB15 data set

Model	*Accuracy*/%	*Precision*/%	*Recall*/%	*F-Measure*/%	Training data set	Testing data set
AlertNet[31]	78.40	94.40	72.50	82.00	U-train	U-test
BCNN[35]	90.25	91.00	90.00	90.45	U-train	U-test
BCNN-DFS[35]	89.26	89.00	89.00	89.00	U-train	U-test
BDNN[36]	80.63	86.00	81.00	79.00	U-train	U-test
NaiveBayes	77.13	83.59	72.74	82.69	U-train	U-test
J48	76.95	70.50	99.98	82.69	U-train	U-test
Random Forest	80.94	74.34	99.84	85.23	U-train	U-test
Bagging	76.95	70.50	99.98	82.69	U-train	U-test
Adaboost	78.13	71.63	99.82	83.41	U-train	U-test
SEMI-GRU	88.11	96.29	81.56	88.13	U-train	U-test

4.4.3 CIC-Bell-DNS-EXF-2021 数据集结果

本文对 CIC-Bell-DNS-EXF-2021 数据集轻文件攻击过程产生的 pcap 文件进行处理，使用从 pcap 文件中构造出的轻文件攻击数据集进行多分类和二分类网络流量异常检测。将 GRU 的层数设置为 3 层，初始学习率设置为 0.001，批大小设置为 256。可以得到 SEMI-GRU 方法对 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击多分类的识别性能，如表 13 所列。

表 13 SEMI-GRU 方法对 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击多分类的识别性能

Table 13 Light file attack multi-category recognition performance of SEMI-GRU method on CIC-Bell_DNS-EXF-2021 dataset

类别	精确率/%	召回率/%	F1 分数/%	支持度
Benign	99.91	99.97	99.94	3519
Light_exe	99.14	98.42	98.78	1645
Light_compressed	98.60	99.40	99.00	2330
Light_video	97.36	98.86	98.11	1231
Light_text	97.43	95.23	96.32	797
Light_audio	99.39	99.54	99.47	1976
Light_image	100.00	91.39	95.50	151

从表 13 可以看出，SEMI-GRU 方法对 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击流量多类别识别效果较好，对 Benign，light_compressed，light_audio 等类别流量识别的 F1 分数均接近于 1。

我们将 Benign 类别作为一个类别，其他的 6 种攻击类型合并在一起作为一个攻击类别，可以得到 SEMI-GRU 方法关于 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击二分类的识别性能，如表 14 所列。

表 14 SEMI-GRU 方法对 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击二分类的识别性能

Table 14 Light file attack two-category recognition performance of SEMI-GRU method on CIC-Bell_DNS-EXF-2021 dataset

类别	精确率/%	召回率/%	F1 分数/%	支持度
Benign	100.00	100.00	100.00	3519
Attack	100.00	100.00	100.00	8130

从表 14 可以看出，SEMI-GRU 方法对 CIC-Bell_DNS-EXF-2021 数据集轻文件攻击二分类的识别效果非常好，测试集中的所有样本均识别正确，因此可以得出，本文提出的 SEMI-GRU 方法进行现实世界网络流量识别的效果也非常好。

5 讨论

本文主要研究如何提高异常流量检测的效果，并尽可能地降低误报率。为了验证所提方法的有效性，我们同时使用 NSL-KDD 数据集和 UNSW-NB15 数据集对其进行了检验。在数据预处理阶段，采用相同的数据预处理方式，将每个数据样本都映射为一个 19 * 19 的二维灰度图。由于这两个数据集都存在类不平衡问题，因此我们采用 SMOTE 算法来对少数类进行过采样处理。但实验发现，使用 SMOTE 算法对训练集进行过采样并不能提升所有数据集的效果，例如它可以提升对 NSL-KDD 数据集进行异常流量检测的效果，但是对 UNSW-NB15 数据集却毫无效果。其原因可能是 UNSW-NB15 数据集的训练集和测试集的数据分布不同，导致在训练集上良好的效果不能迁移到测试集上。我们还发现，SEMI-GRU 的误报率也随着 GRU 层数的增加而增加，但是 GRU 层数达到某个值之后，误报率就会下降。例如在进行 NSL-KDD 数据集多分类异常检测时，误报率在 GRU 网络层数增加到 8 时开始下降；在进行 NSL-KDD 数据集二分类检测时，6 层的 GRU 网络的误报率还不及 5 层的 GRU 网络的误报率的一半。此外，训练时设置的批大小可能会影响 SEMI-GRU 的效果。

我们在对 UNSW-NB15 数据集进行十分类异常流量检测时发现，不同的学习率也会对 SEMI-GRU 的结果造成影响。我们把 GRU 网络的层数固定为 4，然后采取 3 种不同的学习率：0.002，0.02 和 0.01。实验发现，学习率设置为 0.01 时可以取得较好的效果。我们将该效果与其他方法的效果进行比较，如表 11 所列，所提模型在准确率、召回率和 F 值等指标上的表现均优于对比方法。如表 15 所列，本文方法对 UNSW-NB15 数据集中的 Fuzzers，DoS，Analysis，Backdoor，Shellcode 和 Worms 这 6 个类型的攻击流量识别的效果相对较差，但其对 Benign 类型流量识别的效果较好，并且误报率很低，这对于入侵检测系统来说是一个很关键的优势。

实验发现，SEMI-GRU 模型在进行 UNSW-NB15 数据集二分类异常检测时对 Benign 类型流量的检测精确率达到了 99.54%，但召回率却只有 57.6%；进行十分类异常检测时对 Benign 类型流量检测的精确率为 83.63%，召回率达到了 96.92%。因此本文将训练好的二分类和十分类模型相结合进行 UNSW-NB15 数据集的二分类异常流量检测。

实验结果表明,本文方法对于 UNSW-NB15 数据集二分类也可以实现较好的效果。

表 15 SEMI-GRU 方法关于 UNSW-NB15 数据集的十分类混淆矩阵

Table 15 Ten-category confusion matrix of SEMI-GRU method on UNSW-NB15 data set

		Predicted									
		Benign	Generic	Exploits	Fuzzers	DoS	Reconnaissance	Analysis	Backdoor	Shellcode	Worms
Actual	Benign	35860	1	892	2	0	75	170	0	0	0
	Generic	238	18165	461	0	0	7	0	0	0	0
	Exploits	1306	6	9749	0	0	71	0	0	0	0
	Fuzzers	4592	17	1419	5	0	29	0	0	0	0
	DoS	409	23	3634	0	0	23	0	0	0	0
	Reconnaissance	171	7	750	0	0	2568	0	0	0	0
	Analysis	7	0	670	0	0	0	0	0	0	0
	Backdoor	36	0	546	0	0	1	0	0	0	0
	Shellcode	253	0	64	0	0	61	0	0	0	0
	Worms	7	0	36	0	0	1	0	0	0	0

在本文的研究中,使用 GRU 模型作为底层的神经网络,用于从流量数据中提取特征;通过将 GRU 模型设计为双向和多层,从而形成了 MLB-GRU 模型;将 MLB-GRU 模型提取到的特征向量输入所提出的 SR-FNN 中,所并以半监督学习方式进行模型的训练。我们从不同的方面对超参数进行调整来提升 SEMI-GRU 的效果,充分利用了深度神经网络的特征提取和处理能力。在本文提出的 SEMI-GRU 方法中,所设计的 SR-FNN 由 9 个全连接层组成,通过控制不同全连接层的输出维度,实现两次状态压缩和恢复,最后将输出向量的维度压缩为类别的个数,取得了较好的效果。本文还提出了一种 MixMatch 方案的简化版本,从而能够使用半监督学习方式进行模型训练。相较于有监督学习训练方式,半监督训练方式不仅使用训练集数据,还会使用测试集数据,这提高了模型的泛化能力和效果。实验结果表明,本文提出的方法在和 Naive Bayes,J48,Random Fores,Bagging 和 Adaboost 等传统的机器学习方法的比较中表现出了突出的效果,在和对比方法 AlertNet[27],BCNN[31],BCNN-DFS[31],BDNN[32],Improved autoencoder[33] 和 CATBoost[34] 等的比较中,也取得了最好的综合效果。

结束语 本文提出了 SEMI-GRU 方法用于网络流量异常检测,在不同的数据集中的表现相比对比方法均取得了最好的综合效果,并且实现了较低的误报率,可以将其应用于实际的网络系统中检测异常流量。

SEMI-GRU 的优势在于其在训练时能同时利用训练集和测试集数据,具有高效的数据预处理能力,能够有效地从原始数据中提取特征以及可实现较低的误报率。在下一步工作中,我们将对 SEMI-GRU 做进一步的改进。我们将使用无监督学习的方式对原始数据进行维度缩减,这有助于降低无关数据特征给训练带来的干扰,从而提升 SEMI-GRU 的效果。

参 考 文 献

[1] XIAO X,ZHANG S,MERCALDO F,et al. Android malware detection based on system call sequences and LSTM[J]. Multimedia Tools and Applications,2019,78(4):3979-3999.

[2] BALAKRISHNAN S M,SANGAIAH A K. MIFIM—Middleware solution for service centric anomaly in future Internet models[J]. Future Generation Computer Systems,2017,74:349-365.

[3] CREECH G,HU J. A semantic approach to host-based intrusion detection systems using contiguousand discontiguous system call patterns[J]. IEEE Transactions on Computers,2013,63(4):807-819.

[4] LEE W,STOLFO S J,MOK K W. A data mining framework for building intrusion detection models[C]// Proceedings of the 1999 IEEE Symposium on Security and Privacy (Cat. No. 99CB36344). IEEE,1999:120-132.

[5] KHRAISAT A,GONDAL I,VAMPLEW P. An anomaly intrusion detection system using C5 decision tree classifier[C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham:Springer,2018:149-155.

[6] BUTUN I,MORGERA S D,SANKAR R. A survey of intrusion detection systems in wireless sensor networks[J]. IEEE Communications Surveys & Tutorials,2013,16(1):266-282.

[7] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. Yolov4:Optimal speed and accuracy of object detection[J]. arXiv: 2004.10934,2020.

[8] SONG K,TAN X,QIN T,et al. Mpnet:Masked andpermuted pre-training for language understanding[J]. arXiv:2004.09297,2020.

[9] FU Y,LOU F,MENG F,et al. An intelligent network attack detection method based on rnn[C]// 2018 IEEE Third International Conference on Data Science in Cyberspace(DSC). IEEE,2018:483-489.

[10] IMRANA Y,XIANG Y,ALI L,et al. A bidirectional LSTM deep learning approach for intrusion detection[J]. Expert Systems with Applications,2021,185:115524.

[11] CHUNG J,GULCEHRE C,CHO K H,et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv:1412.3555,2014.

[12] BERTHELOT D,CARLINI N,GOODFELLOW I,et al. Mixmatch:A holistic approach to semi-supervised learning[J]. arXiv:1905.02249,2019.

[13] CHAWLA N V,BOWYER K W,HALL L O,et al. SMOTE:synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research,2002,16:321-357.

[14] MOUSTAFA N,SLAY J. UNSW-NB15:a comprehensive data set for network intrusion detection systems(UNSW-NB15 network data set)[C]// 2015 Military Communications and Infor-

mation Systems Conference(MilCIS). IEEE, 2015: 1-6.
[15] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]// IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE, 2009: 1-6.
[16] SAMANEH M, AMGAD H S, PRINCY V, et al. Lightweight Hybrid Detection of Data Exfiltration using DNS based on Machine Learning[C]// The 11th IEEE International Conference on Communication and Network Security(ICCNS). 2021: 3-5.
[17] SCHÖLKOPF B, PLATT J C, SHAWE-TAYLOR J, et al. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443-1471.
[18] ESKIN E, ARNOLD A, PRERAU M, et al. A geometric framework for unsupervised anomaly detection[M]// Applications of Data Mining in Computer Security. Boston: Springer, 2002: 77-101.
[19] SMITH R, BIVENS A, EMBRECHTS M, et al. Clustering approaches for anomaly based intrusion detection[J]. Proceedings of Intelligent Engineering Systems Through Artificial Neural Networks, 2002, 12(1): 579-584.
[20] ERFANI S M, RAJASEGARAR S, KARUNASEKERA S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning[J]. Pattern Recognition, 2016, 58: 121-134.
[21] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special Lecture on IE, 2015, 2(1): 1-18.
[22] BEGGEL L, PFEIFFER M, BISCHL B. Robust anomaly detection in images using adversarial autoencoders[J]. arXiv: 1901.06355, 2019.
[23] ZENATI H, ROMAIN M, FOO C S, et al. Adversarially learned anomaly detection[C]// 2018 IEEE International Conference on Data Mining(ICDM). IEEE, 2018: 727-736.
[24] RADFORD B J, APOLONIO L M, TRIAS A J, et al. Network traffic anomaly detection using recurrent neural networks[J]. arXiv: 1803.10769, 2018.
[25] WANG W, SHENG Y, WANG J, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection[J]. IEEE access, 2017, 6: 1792-1806.
[26] WANG W, ZHU M, ZENG X, et al. Malware traffic classification using convolutional neural network for representation learning[C]// 17 International Conference on Information Networking(ICOIN). IEEE, 2017: 712-717.
[27] VINAYAKUMAR R, ALAZAB M, SOMAN K P, et al. Deep learning approach for intelligent intrusion detection system[J]. IEEE Access, 2019, 7: 41525-41550.
[28] JAVAID A, NIYAZ Q, SUN W, et al. A deep learning approach for network intrusion detection system[C]// Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies(formerly BIONETICS). 2016: 21-26.
[29] INGRE B, YADAV A. Performance analysis of NSL-KDD dataset using ANN[C]// 15 International Conference on Signal Processing and Communication Engineering Systems. IEEE, 2015: 92-96.
[30] WU K, CHEN Z, LI W. A novel intrusion detection model for a massive network using convolutional neural networks[J]. IEEE Access, 2018, 6: 50850-50859.
[31] AL-TURAIKI I, ALTWAIJRY N. A Convolutional Neural Network for Improved Anomaly-Based Network Intrusion Detection [J]. Big Data, 2021, 9(3): 233-252.
[32] ALTWAIJRY N, ALQAHTANI A, ALTURAIKI I. A deep learning approach for anomaly-based network intrusion detection[C]// International Conference on Big Data and Security. Singapore: Springer, 2019: 603-615.
[33] XU W, JANG-JACCARD J, SINGH A, et al. Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset[J]. IEEE Access, 2021, 9: 140136-140146.
[34] RAJ S, JAIN M, CHOUKSEY P. A Network Intrusion Detection System Based on Categorical Boosting Technique using NSL-KDD[J]. IJCNS, 2021, 1(2): 2582-9238.
[35] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv: 1710.09412, 2017.

LI Haitao, born in 1994, postgraduate. His main research interests include cyber security and intrusion detection.

WANG Ruimin, born in 1982, Ph.D, associate professor. Her main research interests include cyber security and IoT device identification.

(责任编辑:何杨)