ISSN：3078-8994
EISSN：3104-4492

基于卷积神经网络的互联网流量异常检测研究

陈心怡

辽宁工程技术大学理学院，辽宁 阜新 123000

摘　要：随着蓝牙技术的广泛应用，其网络流量日益增大，安全问题日益凸显。传统的异常检测方法在处理复杂、高维网络数据时存在效率低、依赖人工特征提取等局限性。本文提出一种基于卷积神经网络（CNN）的蓝牙网络异常检测模型。首先，利用随机森林算法对网络流量特征进行重要性评估与筛选，排除 IP 地址、端口等易误导模型的属性；随后，基于公开数据集 ISCX-IDS-2012 进行数据预处理，将原始流量划分为统一长度的数据流；最后，构建包含四个卷积层、一个池化层和一个全连接层的 CNN 模型进行训练与验证。实验结果表明，该模型在准确率和损失率方面均随迭代趋于稳定，验证了 CNN 在蓝牙网络异常检测中的有效性与可行性。

关键词：网络异常检测；蓝牙网络；卷积神经网络

1 引言

蓝牙凭借简易操作获得认可，节能型数据传输手段，是日常生活中实用的技术手段，和有网络依赖的传输手段相区别，借助蓝牙技术，若干设备可组成临时无线网络，若存在控制数据流量的需要，即便身处网络服务中断的场所，皆可传递文件，伴随蓝牙连接的大规模覆盖，通信流量稳步上升，威胁态势随之形成，既有异常检测方法中，普遍采用时间序列分析手段，神经网络及其衍生方法。

采用神经网络架构的卷积神经网络模型，在视觉分类、语音转写及文本挖掘等多个维度体现显著价值，现已拓展至异常检测场景，该网络架构可实现特征的自动提取与降维处理，实施多层卷积与池化组合，具备数据多层次特征的挖掘能力，从而在网络异常检测中显示出强大的实用性。

2 相关理论

2.1 卷积神经网络概述

卷积神经网络是一类结构较深的神经网络，其核心在于卷积运算。鉴于此模型在处理图像输入时表现优异，研究者可将图像直接作为输入，从而规避了传统识别算法中繁冗的人工特征提取与数据管理流程。通过端到端的网络训练，便能完成基础的异常检测。其典型架构包含输入层、卷积层、池化层、激活函数层以及全连接层。

2.2 蓝牙网络异常检测技术概述

网络攻击主要分为三类：未经授权窃取信息、利用漏洞获取系统权限，以及通过大量异常连接消耗资源的拒绝服务攻击。网络异常检测旨在通过技术手段提前发现这些异常行为，以保障网络安全。

在检测方法上，机器学习通过建立正常行为基准识别异常，主要包括统计模型、监督分类模型和无监督模型等。其中，监督模型依赖标记数据进行训练，而无监督模型可直接在线学习，无需大量离线训练数据。

近年来，深度学习被广泛应用于异常检测，典型模型包括全连接神经网络（FNN）、卷积神经网络（CNN）及循环神经网络（RNN）等。相比传统方法，深度学习凭借其非线性处理能力，在海量数据环境中具有更高检测效率；同时具备更强的自主学习能力，减少了人工干预需求，使检测过程更为高效和智能。

3 异常检测模型和方案设计

3.1 特征选择

3.1.1 随机森林

本文实验采用随机森林模型，借助随机森林（RF）筛选显著特征，随机森林由多棵决策树构成，每棵决策树的创建阶段，通过 Bootstrap 自助采样方法从原始数据集中进行随机抽取，之后综合这些决策树的预测，最终经筛选得出相应结论，该模型对极端值和干扰因素表现出较强的抗干扰能力，面对数据量庞大的情形时，可于数据分析阶段即时获取特征重要性指标 (variable importance measures, VIM)。

运用随机森林对流量特征做重要性分析，即在特征分裂的节点，决策结点分裂时的 Gini 改善量，就样本集包含的每个特征量开展计算，采用特征基尼指数差与全部特征差之和的比值，求得归一化处理后的特征权重，得出特征间的权重分布，对各特征权重进行重要性对比与排序。

3.1.2 重要度评分

实施重要性权重核算时，权重计算过程中不应涉及流 ID、源 / 目的 IP 与端口、协议、时间戳以及外部 IP 这八项特征，即便传统方案采用了这些机制，攻击者或刻意避开常用端口以

ISSN：3078-8994
EISSN：3104-4492

逃避管控及系统规则限制，或是借助生成的虚假 IP 地址躲避追踪，大量端口采用动态分配机制，数据传输过程中存在多应用复用端口现象，模型训练若依据端口号易引发数据误判。

基于这一前提，区分攻击属性主次时，去除 IP 地址、端口号以及时间戳等无关特征的干扰，采用通用性强且稳定的特征定义攻击行为效果更佳，因为数据的统计特征（如包长、流速等）能更稳定、有效地反映攻击行为的本质。

基于运算结果，各分析特征（后向流量包长度均值，数据包长度极值，流持续时长，前向报文平均长度，流速等成分的优先级得分见图 1：

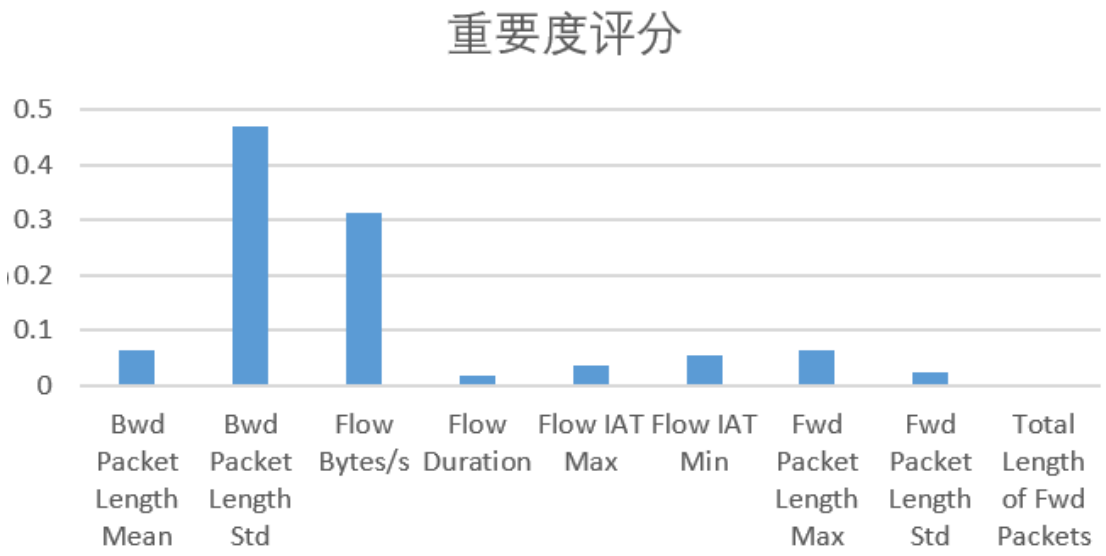

图 1 重要度评分

3.2 数据处理

3.2.1 数据集的选择

本文选择以公开数据集 ISCX-IDS-2012 为实验数据集，此数据集来自加拿大的一个网络安全研究室，ISCX-IDS-2012 入侵检测评估数据集包含以下 7 天的网络活动（正常和恶意），如表 1 所示：

表 1 数据集的网络活动

攻击类型	包大小（G）
正常活动，无恶意活动	16.1
正常活动，无恶意活动	4.22
从内部渗透网络 + 正常活动	3.95
HTTP 拒绝服务 + 正常活动	6.85
使用 IRC 僵尸网络的分布式拒绝服务	23.4
正常活动 . 无恶意活动	17.6
暴力 SSH+ 正常活动	12.3

由于它是使用真实设备构建的，所以创建了真实的正常和恶意流，包括 FTP，SMTP，HTTP，IMAP，SSH 等协议，且所有数据均已被标记。其中攻击种类繁多，包括渗透，拒绝服务，分布式拒绝服务和暴力 SSH，更真实的模拟了被攻击的环境，有利于真实环境的模拟，对现实的网络攻击环境更有参考和实用意义。

3.2.2 数据预处理

从下载的 pcap 文件数据集中，通过 Wireshark 可查看流量详情。由于网络流量数据高维，直接处理开销大，需进行预处理。首先，根据五元组（如 IP 和 MAC 地址）将数据包划分为数据流，每个数据流提取前 125 字节，并仅保留 4 个数据包；若数据包超过 4 个，则将该数据流分割成新流。随后，每种攻击类型的流量分别保存到 Excel 中。为消除无关特征影响（如 MAC 和 IP 地址），进行数据清理，去除这些特征，并对结果进行长度统一，以便后续使用随机森林算法筛选重要特征，并适配卷积神经网络处理。

3.3 网络异常检测模型

3.3.1 实验环境的搭建

本文使用的硬件环境：采用 Win10 操作系统，采用 Python 3.7 环境下的 PyTorch 深度学习平台，借助 PyTorch 功能完成模型拓扑设计及训练样本优化，实验样本的检验，PyTorch 作为 Python 生态中实现深度学习的核心工具库，初始操作应为 Python 环境集成 PyTorch 组件，采用 Anaconda 捆绑的 Python 环境，模块安装就绪后便可立即投入实验。

3.3.2 异常检测模型

基于卷积神经网络的蓝牙异常检测方案包括数据输入、数据预处理、卷积神经网络模型介绍及分类等模块和流程，具体如图 2 所示。

ISSN：3078-8994
EISSN：3104-4492

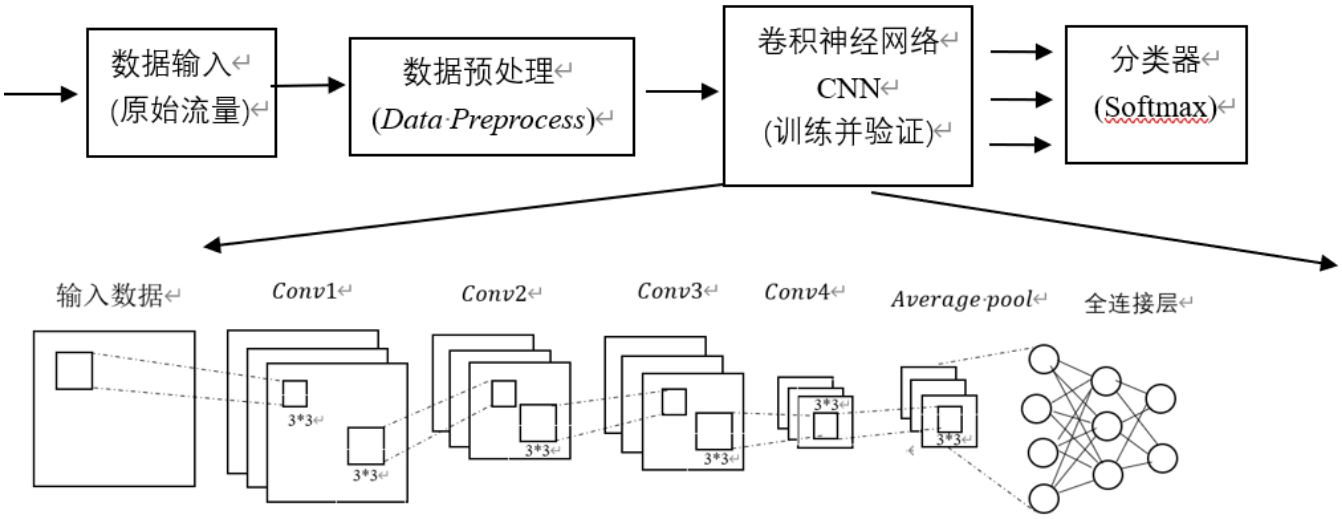

图 2 基于卷积神经网络的蓝牙异常检测方案

从图中可见，系统整合了四个卷积处理阶段，由均值池化层配合 Linear 全连接层组成，针对卷积层的配置阶段，四个卷积层统一采用 3×3 核结构，第三卷积模块采用 2 像素的滑动步长，特征图采用单像素宽度的填充，第四层卷积采用单位步长 stride=1，特征图采用单位宽度边界填充，经过单次卷积运算后，输出矩阵大小为 16 乘以 16，经过第二层卷积处理后，得到 8×8 规格的特征图，实施三层卷积变换后，得到的特征图依然是 8×8 大小；网络最终卷积输出阶段，得到 4×4 大小的输出特征图。平均池化层会得到整个窗口的平均值，可以消除局部极大或极小值带来的影响，使结果会更加稳定。全连接层（或称稠密层）的核心运算是线性变换 y = Wx + b。该运算本身结构简单。然而，为了赋予网络非线性建模能力，每个全连接层之后通常会连接一个非线性激活函数。正是这种线性与非线性操作的交替堆叠，构成了深度神经网络强大的学习基础。

3.4 模型验证

将预处理后的数据集按 1:1 比例进行分割，规范化后的数据平均分配至训练集与实验集，数据准备工作就绪后，采用训练集样本对卷积神经网络实施模型训练，采用 PyTorch 作为 Python 的深度学习架构，进而通过实验数据集检验模型的可靠性，以检验模型的实用价值，进而输出训练精度及损失数值，实验结果显示，随着训练轮次增多，准确度与损失值最终趋于平稳，如图 3 呈现，准确率（Accuracy）伴随迭代次数累积趋向稳定。

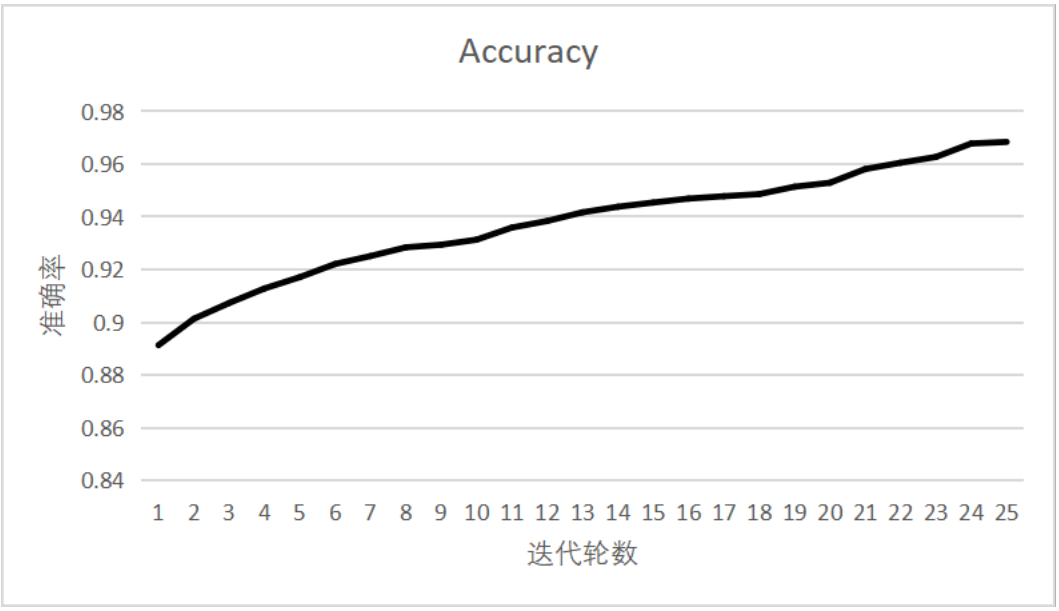

图 3 准确率随迭代轮数变化图

如图 4，表示随着迭代轮数的上升，数据损失率也会下降至平稳。

ISSN：3078-8994
EISSN：3104-4492

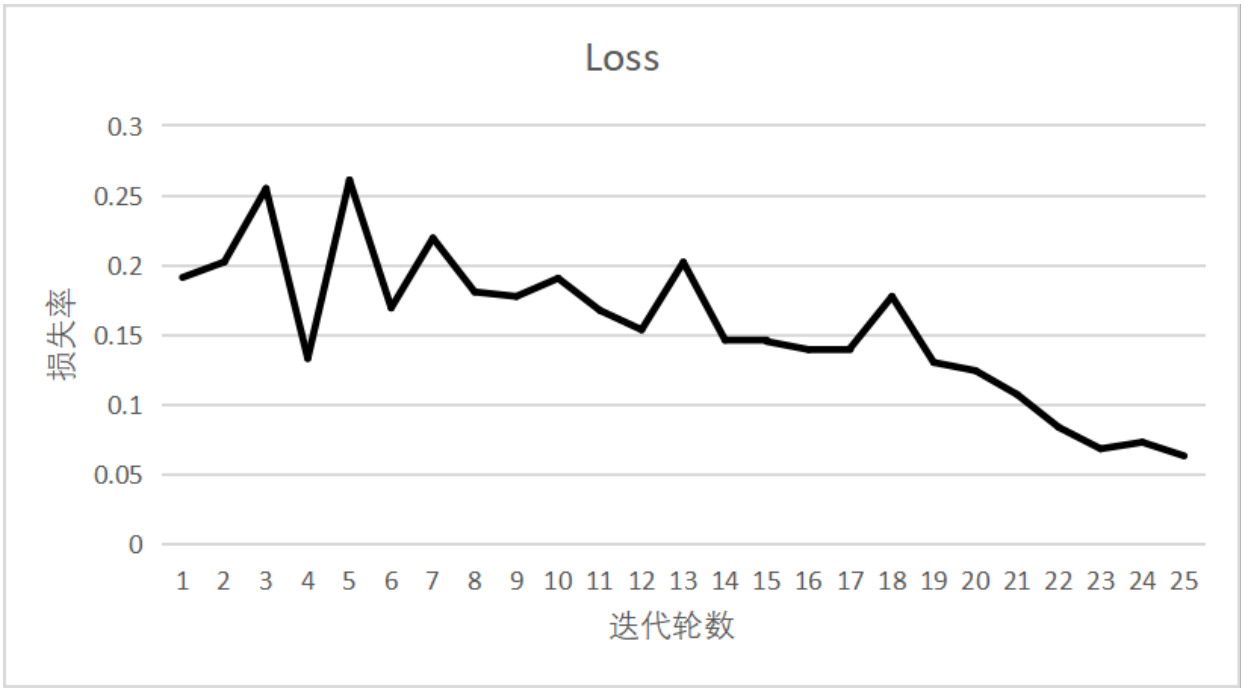

图 4 损失率随迭代轮数变化图

4 结论

本文针对蓝牙网络中的异常检测问题，设计并实现了一种基于卷积神经网络的检测模型。通过随机森林算法筛选关键特征，并结合 CNN 自动提取高维流量特征的能力，有效提升了检测的准确性与效率。实验采用公开数据集进行训练与验证，结果表明模型在处理复杂网络流量时具有良好的稳定性和识别能力。该方法不仅减少了对人工特征工程的依赖，还为蓝牙网络安全提供了一种智能化的检测思路。未来研究可进一步优化模型结构，引入更多真实场景数据，以增强模型的泛化能力与实用性。

参考文献：

[1] 吴迪锋，孙昊翔，曹浪等．网络流量异常检测综述 [J]. 信息安全与通信保密，2022，(08):101-111.
[2] 刘海燕，丛菲．基于深度学习的网络入侵异常检测综述 [J]. 信息系统工程，2020，(09):50-51.
[3] 陈卓，吕娜．基于随机森林和 XGBoost 的网络入侵检测模型 [J]. 信号处理，2020,36(07):1055-1064.
[4] 李光华，李俊清，张亮等．一种融合蚁群算法和随机森林的特征选择方法 [J]. 计算机科学，2019,46(S2):212-215.
[5] 邓华伟，李喜旺．基于深度学习的网络流量异常识别与检测 [J]. 计算机系统应用， 2023，32(2): 274-280.

作者简介：陈心怡（2005.08—），女，汉族，河北省保定市，本科在读，研究方向：机器学习。