人工智能辅助的网络异常行为识别方法

肖博元

（南宁轨道交通运营有限公司，南宁 530000）

摘　要： 网络异常行为识别是保障网络安全的重要方法。文章提出了一种结合长短时记忆网络（Long Short-Term Memory，LSTM）和贝叶斯分类器的网络流量分析方法，旨在通过自动提取网络流量的时间序列特征来提升异常流量识别的准确性。具体而言，首先利用LSTM模型从网络流量中提取隐藏特征表示，随后结合贝叶斯分类器进行流量分类，最后采用NSL-KDD数据集对所提方法进行测试，并通过多项评估。实验结果表明，该方法在识别不同类型流量时均表现出良好的效果，验证了其在网络异常检测领域的实用性和鲁棒性。

关键词： 人工智能；网络安全；流量分析；贝叶斯分类器

中图分类号： TP18　　**文献标识码：** A

Artificial intelligence assisted network abnormal behavior recognition method

XIAO Boyuan

（Nanning Rail Transit Co., Ltd., Nanning 530000, China）

Abstract: Identifying abnormal network behavior is an important method for ensuring network security. The article proposes a network traffic analysis method that combines Long Short-Term Memory (LSTM) and Bayesian classifier, aiming to improve the accuracy of abnormal traffic recognition by automatically extracting time series features of network traffic. Specifically, the LSTM model is first used to extract hidden feature representations from network traffic, followed by traffic classification using a Bayesian classifier. Finally, the proposed method is tested using the NSL-KDD dataset and evaluated through multiple criteria. The experimental results show that the method performs well in identifying different types of traffic, verifying its practicality and robustness in the field of network anomaly detection.

Key word: artificial intelligence, network security, traffic analysis, bayesian classifier

1　引言

随着互联网的快速发展，网络异常行为识别成为保障网络安全的重要方法。这些异常行为通常包括网络攻击、未授权访问、数据泄露等，可能表现为异常的网络流量、异常的用户行为或系统内部的异常活动。

网络异常行为识别方法主要分为基于规则[1]、基于统计[2]和基于机器学习[3]的3大类。基于规则的方法依赖于专家知识，预先定义各种攻击特征并据此进行匹配和检测，优点是实现相对简单且对已知攻击的识别效果较好，缺点在于无法有效应对未知攻击和复杂攻击。基于统计的方法则通过建立网络流量的正常模型来检测与模型显著偏离的行为，对异常行为的检测较为敏感，但在面对变化频繁的网络环境时易产生误报和漏报。基于机器学习的方法主要从大量数据中自动学习特征和模式，无需依赖专家知识。因此，在网络异常行为识别研究中，基于机器学习的方法得到了越发广泛的关注[4]。其中，贝叶斯分类器[5~6]是一种经典的概率模型，在处理不确定性问题时表现出极大的优势。然而，基于传统贝叶斯分类器的网络异常行为识别方法在处理大规模网络流量数据时存在计算复杂度高、实时性差等问题。

综合现有研究，本文首先研究了贝叶斯分类器的基本原理，然后提出了一种基于贝叶斯分类器的网络流量分析与优化方法，最后选用广泛应用于网络安全研究领域的NSL-KDD数据集[7]进行了测试。实验结果表明，该方法在网络异常行为识别方面具有出色的表现。

2　贝叶斯分类器的优化研究

针对传统贝叶斯的问题，本文将基于贝叶斯分类器的异常流量分析方法与自动特征提取技术相结合，以提高分类器的准确性与效率，为网络异常行为识别提供了

一种新的思路和手段。

2.1　贝叶斯分类器的基本原理

贝叶斯分类器是一种基于贝叶斯定理的概率分类方法，其核心思想在于通过计算不同特征与类别之间的条件概率来进行最优分类决策，基本原理可概括为以下步骤：假设有一个样本集 $D=\{x_1,x_2,\cdots,x_n\}$，其中每个样本 x_i 包含多个特征 $X=\{x_{i1},x_{i2},\cdots,x_{im}\}$，需要将其分类为某一类别 $C_j\in\{C_1,C_2,\cdots,C_k\}$ 中的一个。贝叶斯分类器的目标是计算每个类别的后验概率 $P(C_j|X)$，并选择后验概率最大的类别作为样本的预测类别。根据贝叶斯定理，后验概率 $P(C_j|X)$ 可以表示为：

$$P(C_j|X)=\frac{P(X|C_j)\cdot P(C_j)}{P(X)} \tag{1}$$

其中，$P(C_j|X)$ 表示在给定特征 X 的条件下样本属于类别 C_j 的后验概率；$P(X|C_j)$ 表示在类别 C_j 条件下观察到特征 X 的似然函数；$P(C_j)$ 是类别 C_j 的先验概率，反映了类别 C_j 出现的可能性；$P(X)$ 是特征 X 的边际概率，对所有类别的后验概率进行归一化。

为了简化计算，通常假设各特征之间相互独立，即独立性假设。基于此，似然函数 $P(X|C_j)$ 可以分解为各特征条件概率的乘积：

$$P(X|C_j)=\prod_{i=1}^{m}P(x_i|C_j) \tag{2}$$

其中，$P(X|C_j)$ 是在类别 C_j 条件下第 i 个特征 x_i 的条件概率。

基于上述公式，可以计算样本属于每个类别的后验概率，并选择最大后验概率对应的类别 C_{MAP} 作为预测结果：

$$C_{\text{MAP}}=\arg\max_{C_j}P(C_j|X) \tag{3}$$

通过上述推导，贝叶斯分类器能够有效利用先验知识和观测数据计算出样本属于各个类别的概率，从而实现对网络流量的分类。

2.2　特征自动提取与异常流量分析

为了优化基于贝叶斯分类器的网络流量分析方法，本文采用了一种基于 LSTM[8-9] 的网络流量特征自动提取方法。给定一个网络流量序列 $S=\{s_1,s_2,\cdots,s_T\}$，其中 s_t 表示在时间步 t 的网络流量数据。LSTM 网络通过一系列记忆单元来处理该序列，并输出每个时间步的隐藏状态 h_t 作为特征表示。具体来说，LSTM 的运算可以分为以下步骤。

（1）输入门决定了当前输入 s_t 对当前状态的影响，可表示为：

$$i_t=\sigma(W_i s_t+U_i h_{t-1}+b_i) \tag{4}$$

其中，W_i 是输入到输入门的权重矩阵；U_i 是前一隐藏状态 h_{t-1} 到输入门的权重矩阵；b_i 是偏置项；σ 是 Sigmoid 激活函数。

（2）遗忘门决定前一时刻的记忆单元状态对当前时刻的影响：

$$f_t=\sigma(W_f s_t+U_f h_{t-1}+b_f) \tag{5}$$

其中，W_f 和 U_f 是相应的权重矩阵；b_f 是偏置项。

（3）细胞状态更新结合输入门和遗忘门的信息来更新记忆单元状态：

$$c_t=f_t\cdot c_{t-1}+i_t\cdot\tanh(W_c s_t+U_c h_{t-1}+b_c) \tag{6}$$

其中，W_c 和 U_c 是状态更新的权重矩阵；tanh 是双曲正切激活函数。

（4）输出门的输出 o_t 为：

$$o_t=\sigma(W_o s_t+U_o h_{t-1}+b_o) \tag{7}$$

其中，W_o 和 U_o 是相应的权重矩阵；b_o 是偏置项。

（5）输出隐藏状态 h_t：

$$h_t=o_t\cdot\tanh(c_t) \tag{8}$$

通过上述过程，LSTM 网络能够从输入的网络流量序列中自动提取出时间步的隐藏状态 $H=\{h_1,h_2,\cdots,h_T\}$，这些隐藏状态可被视作反映网络流量动态变化的高维特征表示。下一步是将这些 LSTM 提取的特征与贝叶斯分类器相结合以进行异常流量分析，步骤如下。

（1）特征提取：利用训练好的 LSTM 网络对网络流量序列 $S=\{s_1,s_2,\cdots,s_T\}$ 进行处理，得到隐藏状态特征 $H=\{h_1,h_2,\cdots,h_T\}$。

（2）特征选择：从提取的特征中选择最能区分正常流量与异常流量的特征子集 $\hat{H}=\{\hat{h}_1,\hat{h}_2,\cdots,\hat{h}_n\}$，以提高分类器的性能。

（3）贝叶斯分类：对于每个特征向量 $\hat{h}_i$，利用贝叶斯分类器计算其属于各个类别的后验概率 $P(C_j|\hat{h}_i)$，其中 C_j 是类别标签（如正常流量或异常流量）。

（4）分类决策：根据后验概率的大小，选择最大后验概率对应的类别 C_{MAP} 作为网络流量的分类结果。

3　实验与分析

本文采用 NSL-KDD 数据集对方法进行了测试。该数据集是网络安全研究领域广泛使用的标准数据集，由正常流量记录和多种类型的攻击流量记录组成，包括拒绝服务攻击（Denial of service，DoS）、用户到根攻击（User to root，U2R）、远程到本地攻击（Remote to local，R2L）以及探测攻击（Probe）等，被用于评估入侵检测系统的性能。为了验证基于 LSTM 的网络流量特征自动提取与贝叶斯分类器相结合的网络流量识别方法的有效性，本文使用 MATLAB 设计[10] 了以下实验方案。

（1）数据预处理：在 MATLAB 中加载 NSL-KDD 数据集的训练集和测试集，并用标签标识该样本为正常流量或异常流量。

（2）基于 LSTM 的特征自动提取：在 MATLAB 中构建一个 LSTM 网络，用于处理输入的时间序列网络流量数

据;利用训练集数据对LSTM网络进行训练,目标是最小化预测特征与真实标签之间的误差;训练完成后,利用训练好的LSTM网络对测试集的流量数据进行处理,提取每个流量样本的隐藏状态作为特征表示。

(3)贝叶斯分类器的分类:在MATLAB中实现贝叶斯分类器,输入为LSTM提取的特征,输出为每个流量样本的类别标签;利用训练集的特征提取结果估计各类别的先验概率$P(C_j)$和条件概率$P(\hat{h}_i|C_j)$。

(4)分类预测:利用贝叶斯分类器对测试集数据进行分类,计算每个样本属于各类别的后验概率$P(C_j|\hat{h}_i)$,并将最大后验概率对应的类别作为预测结果。

为对该方法进行评估,本实验采用了准确率、精确率、召回率和F1值等指标,如表1所列。

表1 实验结果评估

类别	准确率	精确率	召回率	F1值
正常流量	0.98	0.97	0.99	0.98
DoS	0.95	0.96	0.94	0.95
U2R	0.92	0.91	0.90	0.91
R2L	0.89	0.88	0.87	0.88
Probe	0.96	0.95	0.96	0.95

从表1可以看出,本文方法在多个评估指标上均表现出良好的效果。首先,从准确率的角度来看,各类流量的识别准确率均超过了0.89,尤其是在正常流量和DoS攻击流量的识别上,准确率分别达到0.98和0.95,表明该方法能够较为准确地区分正常流量与各类攻击流量。其次,精确率指标显示该方法在不同攻击类型的识别上具有较高的精度,特别是在DoS攻击和Probe上,精确率分别达到了0.96和0.95。这表明该方法能够有效降低误报率,准确检测出异常流量而不误判正常流量,提升了网络安全系统的可靠性。在召回率方面,该方法在正常流量和Probe的识别上具有更好的表现,分别达到了0.99和0.96。最后,从F1值来看,本文方法在各类流量的识别上表现出高度的平衡性,特别是在DoS攻击和探测攻击的识别上,F1值均为0.95,表明其在实际网络环境中能够同时兼顾识别精度与全面性,提供高效的异常流量检测。综上所述,基于LSTM特征提取与贝叶斯分类器结合的方法在网络流量异常检测任务中具有出色的表现,特别是在识别精度和全面性方面具有显著优势。这一结果进一步证明了本文方法在网络安全领域中的实用性和有效性。

4 总结

本文在网络异常行为识别领域引入了基于LSTM的特征自动提取与贝叶斯分类器相结合的分析方法。实验证明,该方法在NSL-KDD数据集上具有良好的表现,特别是在识别精度和全面性方面具有显著优势,进一步证明了该方法在实际应用中的潜力与价值。然而,尽管该方法取得了较好的实验结果,但其在应对更大规模和更加多样化的网络流量时仍可能面临挑战,未来的研究可以考虑结合更先进的深度学习模型或优化贝叶斯分类器的决策过程,以进一步提升识别效果和计算效率。

参考文献:

[1] 张琪鑫,吴超,罗娜,等.基于规则拟合的TCP数据包流量混淆系统[J].计算机应用与软件,2018,35(2):145-149.
[2] 周燕茹.一种基于统计模型的网络阻塞攻击防御方法[J].遵义师范学院学报,2022,24(2):94-98.
[3] 张蕾,崔勇,刘静,等.机器学习在网络空间安全研究中的应用[J].计算机学报,2018,41(9):1943-1975.
[4] 张博,刘绚,于宗超,等.基于人工智能的电力系统网络攻击检测研究综述[J].高电压技术,2022,48(11):4413-4426.
[5] 文志诚,曹春丽,周浩.基于朴素贝叶斯分类器的网络安全态势评估方法[J].计算机应用,2015,35(8):2164-2168.
[6] 胡滨,代昆玉,王翔.改进贝叶斯分类算法在DDoS攻击检测系统中的研究[J].贵州大学学报(自然科学版),2010,27(3):84-87.
[7] 朱平哲.基于NSL-KDD数据集的物联网入侵检测特征选择方法研究[J].江苏工程职业技术学院学报,2019,19(3):17-21.
[8] 杨丽,吴雨茜,王俊丽,等.循环神经网络研究综述[J].计算机应用,2018,38(S2):1-6+26.
[9] 麻文刚,张亚东,郭进.基于LSTM与改进残差网络优化的异常流量检测方法[J].通信学报,2021,42(5):23-40.
[10] 李隆烨.基于MATLAB的贝叶斯分类器设计[J].科技传播,2019,11(20):116-117.

作者简介:

肖博元(1986—),本科,工程师,研究方向:网络安全、人工智能。