DOI: 10.13382/j.jemi.B2407379

# 因素空间背景基的流量异常检测基点分类方法*

陈万志[1] 任鹏江[1] 王天元[2]
(1. 辽宁工程技术大学软件学院 葫芦岛 125105;2. 国网辽宁省电力有限公司 营口 115005)

**摘 要**:针对机器学习在流量异常检测中存在特征选择依赖经验、易受离群点影响导致鲁棒性差等问题,基于因素空间理论的“背景关系-背景分布-背景基”体系提出一种流量异常检测的基点分类方法。首先,数据预处理阶段使用 KNN 离群点检测算法去除数据中的离群点,降低异常点对后续背景基提取的影响。其次,使用 mRMR 算法对数据特征进行排序,选择对分类最具影响力的特征标注为类别区分特征。然后,以内点判别法为理论基础优化背景基提取算法,提取训练数据中不同类别数据的背景基,得到各类别的单位认知包。最后,以单位认知包为核心构造基点分类算法(fundamental point classification algorithm, FPCA)实现异常流量的精准二分类。在 NSL-KDD 数据集上对所提方法的二分类实验准确率和 *F*1-*score* 分别达到 92.48%和 92.18%,检测性能优于同类型的其他机器学习方法。在 CICIDS2017 场景数据集上的测试进一步验证了所提方法在实际应用中的可行性。

**关键词**: 因素空间;背景基;基点分类;异常检测
**中图分类号**: TP393;TN911.7 **文献标识码**: A **国家标准学科分类代码**: 510.40

# Traffic anomaly detection method based on fundamental point classification by factor space background basis

Chen Wanzhi[1] Ren Pengjiang[1] Wang Tianyuan[2]
(1. College of Software, Liaoning Technical University, Huludao 125105, China; 2. State Grid YingkouElectric Power Company of Liaoning Electric Power Supply CO, Yingkou 115005, China)

**Abstract**: In order to solve the problems of feature selection dependent on experience and poor robustness caused by outliers in machine learning traffic anomaly detection, a fundamental point classification method for traffic anomaly detection based on the “background relation-background distribution-background basis” system by factor space theory is proposed. Firstly, the KNN outlier detection algorithm is used to remove outliers in the data in the data preprocessing stage to reduce the influence of outliers on the subsequent background basis extraction. Secondly, the mRMR algorithm is used to sort the data features and select the most influential features for classification as category distinguishing features. Then, the background basis extraction algorithm is optimized based on the internal point discriminant method, and the background basis of different types of data in the training data is extracted, and the unit cognition package of each type is obtained. Finally, a fundamental point classification algorithm (FPCA) based on the unit cognitive packet is constructed to achieve accurate two-class classification of abnormal traffic. The proposed method attains accuracy rate of 92.48% and F1-score of 92.18% in a two-class classification task on the NSL-KDD dataset, which detection performance superior to the same type machine learning method. The test on CICIDS2017 scene data set further verifies the feasibility of the proposed method.

**Keywords**: factor space; background basis; fundamental point classification; anomaly detection

收稿日期: 2023-10-13 Received Date:
*基金项目:国家重点研发计划(2018YFB1403303)、辽宁省教育厅高校科研基金(2021LJKZ0327)项目资助

## 0 引　言

信息技术的飞速发展使得海量网络应用为用户提供便捷服务的同时,针对网络协议和应用程序漏洞的入侵攻击屡屡发生,网络空间的安全风险日益加剧。而流量异常检测可以及时发现潜在的威胁和入侵迹象,是一种解决网络攻击问题行之有效的防护手段。

随着机器学习和深度学习大模型的普及应用,在网络安全领域,利用机器学习技术对入侵检测系统进行相关数据集训练和分析,有效提升入侵检测系统识别异常能力。宋雅洁等[1]利用隐马尔科夫模型并结合网络流量特征分析实现系统的故障信息与攻击数据的区分,从而提升列车运行控制系统的入侵检测能力[2-3]。王琳琳等[4]结合极限学习机(extremal learning machine, ELM)与 K-means 算法的级联分类器模型对各种类型攻击的检测性能较好,但 ELM 算法的参数未进行优化且 K-means 算法对于初始聚类中心选择存在较大程度依赖,因此造成检测效果并不佳。陈万志等[5]提出一种融合 AdaBoost(adaptive boosting, AdaBoost)与 BP(back propagation, BP)神经网络的入侵检测方法,有效避免神经网络易陷入局部最优的现象,但未知攻击类型的检测能力有限。陈万志等[6]采用白名单技术对工控网络数据进行一次过滤后,再结合自适应变异粒子群优化的神经网络算法对通信异常行为进行二次过滤,能有效识别未知攻击行为,提升了检测的准确率,但无法保证工业控制网络入侵检测的实时性要求。王华忠等[7]采用粒子群优化(particle swarm optimization, PSO)和支持向量机(support vector machine, SVM)的核函数以及惩罚参数提升了系统的分类性能,但对各类典型攻击数据检测率不高且 PSO 算法长时间迭代寻优易陷入局部最优。梁辰等[8]通过主成分分析法进行数据特征降维以及权值修正,对 BP 神经网络算法加以改进使得模型检测效率得到有效提升,但算法参数因工控网络数据具体情况变化导致算法不稳定。

深度学习具备出色的表征学习能力,能够从原始数据中自主发掘并提取关键特征,实现精准的分类判断。常用方法包括堆叠降噪自动编码器(stacked denoising auto-encoder, SDA)[9]、受限玻尔兹曼机(restricted Boltzmann machine, RBM)[10]、卷积神经网络(convolutional neural network, CNN)[11]等。吕佩吾等[12]应用 HOOK 机制将 CNN 模型嵌入工业防火墙实现 Modbus/TCP 的异常报文检测,所提检测模型能够有效解决基于 Modbus/TCP 协议产生的相关漏洞问题。李熠等[13]提出一种稀疏自编码器(sparse auto-encoder, SAE)与 ELM 相结合的入侵检测方法,首先利用 SAE 对工控网络的高维非线性数据特征进行集成式特征提取,然后应用 ELM 对抽象特征进行分类,满足工业控制系统入侵检测的检测率及误报率的要求。王竹晓等[14]利用深度强化学习算法对电力工控网络的异常通信行为进行检测,模型采用 Q-learning 与神经网络相结合方法对电力工控网络数据进行训练,可有效检测出电力工控网络异常通信行为。梁欣怡等[15]提出一种基于自监督特征增强的卷积神经和双向长短期记忆网络入侵检测方法,解决了检测中攻击样本和流量特征不足的问题。此外,生成式神经网络在应对流量样本类别分布不均的问题上也发挥了重要作用。其中,生成对抗网络(generative adversarial network, GAN)[16]和自编码(auto-encode, AE)神经网络是两种主要的方法,其通过对正常流量进行建模,能够有效地重构正常流量的特征,而对于异常流量,则在重构过程中会产生显著的偏差,这种偏差为识别和区分异常流量提供一个有效的方式。虽然这些模型在特征网络环境中能够对异常流量进行识别,但它们大多采用单层体系结构,仅利用了流量数据在单一尺度下的特征,在不同尺度下的多样性特征未被充分利用,无法保证同时满足检测精度、误报率以及对不同网络环境的适应度等实际需求。

针对上述存在问题,本文基于因素空间理论的“背景关系-背景分布-背景基”体系[17-19]提出一种新的流量异常检测基点分类方法。从概念与推理两个基本面发现样本数据中因素特征之间的内在联系,通过挖掘流量类别与因素特征之间的耦合关系进行多角度学习得到能够表征各类别样本的背景基,形成对正常流量和攻击流量具有较高识别度的类别认知包,将其作为基点分类算法中识别异常流量的基础知识完成异常检测。

## 1 因素空间理论基础

因素是事物构成的本质,是信息科学区别于物质科学的关键词。区分两个物质的关键就在于其各自的属性,不同事物具有不同的特征和属性。变量是由属性或其他形式的信息所表征的值,这些信息被视为因素。因素比属性更高一个层次,是属性的统领,无是非可言。属性是定词,是某种具体的表现。因素与属性不同地方在于因素不直接表述客观事实,而是在一定条件下,通过某些方式把事物的特征反映出来。事物的属性呈现出一种被动的状态,而因素则更具有启示性和引导性,主动引导着人类的思维。在一定程度上说,因素比属性更能揭示客观现实中的本质联系。下面给出因素的定义。

定义 1　一类事物的集合称为论域,用 $U$ 表示。

定义 2　映射 $f:U\to X(f)$ 称为一个因素,其中 $X(f)$ 是 $f$ 从事物所映射出来的性状集合,称为 $f$ 的性状空间。

对于给定论域 $U$ 上的一组因素:

$f_j: U \to X(f_j), j = 1,2,\cdots,n$

称集合 $F^* = \{f_1, f_2, \cdots, f_n\}$ 为当前论域 $U$ 的一个因素集。

$P(F^*)$ 为 $F^*$ 的幂集，元素个数为 $2^{|F^*|}$，此处 $|F^*| = n$，$P(F^*)$ 的任意元素 $F_i = \{f_{(j)}, \cdots, f_{(k)}\}$，其中 $i = 1,2,\cdots,2^{|F^*|}$，$\{f_1, f_2, \cdots, f_n\}$ 是 $P(F^*)$ 的一个子集合，定义一个 $U$ 上的合成因素 $F_i: U \to X(F_i)$，其性状空间是：

$X(F_i) = X(f_{(j)}) \times \cdots \times X(f_{(k)})$

记为 $F_i = f_{(j)} \cup \cdots \cup f_{(k)}$。其含义是合成因素 $F_i$ 的性状空间 $X(F_i)$ 被定义成一个笛卡尔乘积 $X(f_{(j)}) \times \cdots \times X(f_{(k)})$。

定义 3　因素空间：记 $X_{(F^*)} = \{X(F)\}_{(F \in P(F^*))}$，称 $\boldsymbol{\Phi} = (U, X_{F^*})$ 为 $U$ 上的一个因素空间。有时也可表示为 $\boldsymbol{\Phi} = (U, F^* = \{f_1, \cdots, f_n\})$。

记 $X_{\max} = X(f_1) \times X(f_2) \times \cdots \times X(f_n)$ 称为最大性状空间，对于离散型性状空间而言，任意向量 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n) \in X$ 称为一个性状颗粒。

定义 4　背景关系：给定 $U$ 上的定性因素空间 $\boldsymbol{\Phi} = (U, X_{F^*})$，对任意相 $a = (a_1, a_2, \cdots, a_n) \in X$，记其在 $U$ 上的原相为：

$[a] = F^{-1}(a) = \{u \in U \mid F(u) = a\}$

$[a]$ 可能是空集，若 $[a] \neq \varnothing$，则称 $a$ 是一个实性状颗粒，否则称 $a$ 是一个虚组态。全体实性状集合记为：

$R = F^*(U) = \{a = (a_1, a_2, \cdots, a_n) \in X \mid \exists u \in U$

$f_1(u) = a_1, f_2(u) = a_2, \cdots, f_n(u) = a_n\}$

式中：$R$ 称为因素 $f_1, f_2, \cdots, f_n$ 之间的背景关系，也称为因素 $F^*$ 的背景集。

机器学习中 $a$ 可以视为一条样本，$[a]$ 表示 $a$ 的对应标签。如果一个 $a$ 有对应标签，则 $a$ 是有意义的，即 $a$ 是一个实性状颗粒，否则 $a$ 是没有意义的，称 $a$ 是一个虚组态。背景关系可以理解为：全部有意义的样本构成的集合 $S$ 与样本的标签之间的因果关系。若要深入讨论集合 $S$ 的分布问题，则引出定义 5。

定义 5　背景分布：设论域 $U = (U, A, p)$ 是一个概率场，$\boldsymbol{\Phi} = (U, X_{F^*})$ 是定义在 $U$ 上的一个因素空间，$X = (X, B)$ 是最大性状空间上的一个可测结构。若所有 $F^*$ 中的 $f_j$ 都是可测映射，记 $p = p_{F^*}$ 为 $p$ 经过 $F^*$ 在 $X$ 上所诱导出来的概率，亦即对任意 $B \in B$，都有 $p(B) = p((F^*)^{-1}(B))$。$B$ 称为因素 $F^*$ 的背景分布。

定义 6　背景基：若每个性状空间 $X(f_j)$ 都是有序集，背景关系 $R$ 是 $X$ 中的凸集，记 $R$ 的所有顶点所构成的集合为 $B = B(R) = \{P \mid P$ 是 $R$ 的顶点$\}$，称作背景基。$S$ 表示样本集合，当 $S = R$ 时，记 $B$ 的所有顶点构成的集合为 $B(S) = \{P \mid P$ 是 $S$ 的顶点$\}$，称为样本背景基。

背景基可以生成背景关系，是背景关系的无信息损失的压缩，对因素库的实际应用具有重要的意义。为了更好地说明背景关系、背景分布、背景基三者概念之间的联系，以图 1 为例进行图示说明。

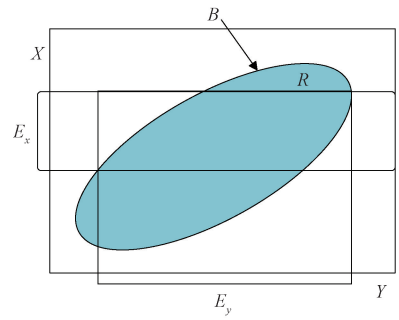


图 1　“背景关系-背景分布-背景基”关系

Fig. 1　“Background relation-background distribution-background basis” relationships

设集合 $X$、$Y$、$R$，其中 $E_x \subseteq X$、$E_y \subseteq Y$，集合 $R$ 称为背景关系，集合 $R$ 的“边缘”构成背景基 $B$，$p$ 是背景分布，视为集合 $R$ 的概率密度，可以由样本的相 $a$ 的频率逼近来实现。背景关系 $R$ 表示集合 $X$ 和集合 $Y$ 之间的关联性。这里有：

$E_x R E_y$

可得逻辑推理

$E_x \to R_y(R: \to)$

因此可以认为，背景关系 $R$ 保证了 $X$ 与 $Y$ 的因果关系，背景基则限定了背景关系区域，背景分布表示了因果关系的概率。

## 2　因素空间背景基的流量异常检测基点分类方法

### 2.1　总体框架

本文提出的因素空间背景基的流量异常检测基点分类方法，包括前期的数据预处理、背景基提取算法的实现以及 FPCA 异常检测 3 个部分，总体框架如图 2 所示[20]。

方法的主要思路如下：

1）对原始数据集进行数据清洗、字符特征数值化、数据归一化等预处理得到样本数据 Sample，并将样本数据 Sample 按照类别划分为多个数据集，对每个数据集去除离群点；2）使用 mRMR 算法对数据特征进行排序，选择对分类最具影响力的特征标注为类别区分特征；3）使用背景基提取算法提取训练数据中不同类别数据的背景基，得到各个类别的背景基点 ***B_class***，得到各个类别的单位认知包；4）以各个单位认知包为核心构造基点分类算法，将测试数据作为各类别认知包的输入，在认知包中

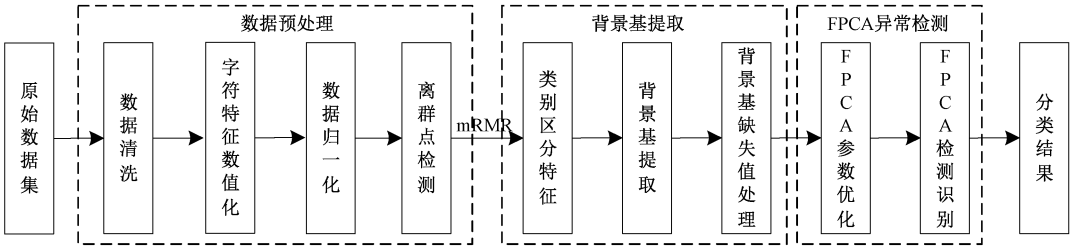


图 2　总体框架

Fig. 2　Overall framework structure

对数据进行识别,若符合某一认知包特征,则将该条数据认定为此类数据,未匹配相应认知包的数据归为未知攻击类型。

## 2.2　数据预处理

1)数据清洗

在流量异常检测的数据集中,往往某些数据没有记录或无法获取,由此产生的缺失值可能会导致异常检测的误判,并且可能存在一些与流量异常检测无关的数据,这些数据也会干扰异常检测的准确性,通过数据清洗,去除缺失值、无关数据等不良数据,可以增强数据的质量,降低异常检测的误判率,提高准确性。

2)字符型特征数值化

流量数据中含有字符型特征,如 NSL-KDD 数据集中的'protocol type'、'service'和'flag',为了满足后续背景基提取的需求,需要将这些字符型特征转换为数值特征,本文采用 LabelEncoder( ) 函数对非数值特征进行标签编码。同时,为了确保模型能够正确地识别样本的类别,样本的类别标签需要被转化为数字形式,在实验测试中,将正常流量数据的标签编码为 0,异常流量数据的标签编码为 1。

3)数据归一化

数据集中特征值间的大小差异可能会对算法产生不良影响,为确保检测结果的准确性并减少不同特征之间量纲差异所带来的影响,本文采用 Min-Max 方法对数据进行归一化处理,将各个特征的数据范围映射到[0,1]区间内,其表达式为:

$$x' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

其中,$X$ 为初始特征值,$X_{\min}$ 和 $X_{\max}$ 分别为对应特征的最小值和最大值。

4)离群点检测

离群点是数据集中与其他数据点明显不同或者极端的数据点,具有与大多数同类别数据点不同的特征或行为。为了减小异常数据点对后续背景基提取和分类过程可能产生的误导和误判,采用 KNN 离群点检测算法来识别和剔除这些异常点。该算法的核心思路是计算数据集中每个点与其 $k$ 个最近邻点之间的平均距离,并对这些距离按照降序排列,将距离最大的前 $n$ 个点视为离群点,从而实现对异常值的检测和清除。

## 2.3　改进的背景基提取算法

因素空间的"背景关系-背景分布-背景基"理论体系追求使用数学的方式模拟人的学习过程。当前技术层面上所采用的基于夹角判别定理的背景基提取算法存在显著的不足:其判别条件复杂,时间复杂度较高,且提取出的基点数量过多,这些因素使得学到的知识在后续的继承和保存时存在困难。

内点判别法具有计算简洁和基点提取效率高的特点,尤其在处理涉及大量样本和高维度数据的场景中,具备卓越的性能和优势,以内点判别法为核心对背景基提取算法进行优化,不仅能够降低算法的复杂度,还能模仿人类的学习模式,勾勒出知识的框架,并高效地保留所习得的知识,更加贴近人的思维逻辑。综上,本文提出改进的背景基提取算法(improved background basis extraction algorithm, IBBEA),具体描述如下。

算法 1:IBBEA 算法

输入:样本集合 Sample
输出:背景基集合 ***B_class***
0: ***B_class*** = ∅,计数器 *Counter*=0,背景基个数 *Number*=15
1: for item in Sample: #遍历样本集合中的元素
2:　if *Counter* < *Number*:
3:　　将样本元素 item 放入 ***B_class***
4:　　*Counter*+= 1
5:　else:
6:　　计算 ***B_class*** 的几何中心 $o$:

$$o = \frac{1}{k}\sum_{j=1}^{k} b_j, b_j \in \boldsymbol{B\_class}, j = 1,2,\cdots, |\boldsymbol{B\_class}|$$

7:　　计算 $(\alpha,\beta_j) = \alpha\beta_j^{\mathrm{T}}$
　　其中 $\alpha = o-\text{item}, \beta_j = b_j-\text{item}, j=1,2,\cdots,|\boldsymbol{B\_class}|$
8:　　if $(\alpha,\beta_j) < 0$:
9:　　continue
10:　else:
11:　　将样本元素 item 放入 ***B_class***
12:　　*Counter*+= 1
13:　　for B_item in ***B_class***:
14:　　if isInnerPoint(B_item, ***B_class***-B_item):
15:　　if B_item == $P_j^+$ or B_item == $P_j^-$ :

```
        # P_j^+ = argmax_i{b_ij | b_ij ∈ B_class}
        # P_j^- = argmin_i{b_ij | b_ij ∈ B_class}
16:     更新背景基集合 B_class:
            B_class\B_item → B_class
17:         Counter-= 1
18:   end for
19: end for
```

IBBEA 算法用于从给定样本集合中提取出可以表达样本集的背景基集合,其详细解释描述如下。

参数声明:初始化一个空集合 ***B_class*** 用于存放提取得到的背景基;初始化计数器 *Counter* 用于记录提取出的背景基数量,初始值设为 0;初始化背景基个数 *Number*,初始值设为 15;用于控制初始阶段添加到背景基集合的样本数量。

步骤 1)循环遍历样本集合中的每个元素。

步骤 2)~4)在每次循环中,首先检查 *Counter* 的值是否小于预定的初始背景基个数 *Number*。如果是,意味着背景基集合尚未达到初始设定的大小,因此直接将当前样本元素添加到 ***B_class*** 中,并递增 *Counter*。

步骤 6)当 *Counter* 的值达到或超过预定的初始背景基个数 *Number* 时,计算当前背景基集合 ***B_class*** 的几何中心 $o$,中心点 $o$ 根据背景基中所有元素的位置计算得到。

步骤 7)计算当前样本元素 *item* 分别与几何中心 $o$ 和其他背景基元素 $b$ 之间的距离,记作$(\alpha, \beta)$。

步骤 8)~12)判断$(\alpha, \beta)$的值是否小于 0,若是,则跳过该样本,继续处理下一个样本元素;否则,将当前样本元素添加到背景基集合 ***B_class*** 中,并递增 *Counter*。

步骤 13)~18)新增背景基元素后,遍历 ***B_class*** 中的元素,通过调用 isInnerPoint( ) 函数判断背景基集合中的每个元素 *B_item* 是否属于更新后的背景基集合的内点,若是,且 *B_item* 为 ***B_class*** 中的极值,则认为 *B_item* 不再是背景基的有效组成部分,将其从 ***B_class*** 中移除,并递减 *Counter*。

步骤 19)结束循环,并返回构建好的背景基集合 ***B_class***。

背景基 ***B_class*** 以矩阵的形式进行存储,若样本 *Sample* 包含 $n$ 个因素,且 ***B_class*** 的基点数量为 $m$,则 ***B_class*** 将呈现为一个 $m\times n$ 的矩阵形式,这样的矩阵结构能够有效地组织和表示背景基中的各个基点及其对应的因素信息为:

$$\boldsymbol{B\_class}_{m\times n}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

式中:***B_class*** 的每一行表示一个基点。

## 2.4 基点分类算法

IBBEA 算法训练得到的模型是相对固定。而在实际中,决策者的主观因素常常扮演重要角色,因此 FPCA 算法在做决策时需要加入决策者主观参数,故由主观参数变换得到的背景基称为 $\lambda$-背景基。

定义 7 设集合 $B_i \neq \varnothing$,由 $c$ 个类别生成的背景基 $B=\bigcup_{i=1}^{c} B_i$,各类别用 $i$ 表示,$i=1,2,\cdots,c, b_j \in B$,则 $[B]_i=\{b_j^i \mid i \in c, j=1,2,\cdots,n\}$,其中 $n=|[B]_i|$,称 $[B]_i$ 为类别 $i$ 的背景基。

定义 8 设 $B$ 为 $c$ 个类别的背景基,$[B]_i$ 为类别 $i$ 的背景基,$b_j \in [B]_i$,向量 $\boldsymbol{\Lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_c)$,$\lambda_i>0$,$i=1,2,\cdots,c$,使得:

$$[B_\lambda]_i=\boldsymbol{\Lambda}\otimes[B]_i=\{\lambda_i B_i \mid j=1,2,\cdots,|[B]_i|\} \tag{2}$$

则称 $B_\lambda=\bigcup_{i=1}^{n}[B_\lambda]_i$ 为 $\lambda$-背景基。

从几何学角度看,$\lambda$-背景基 $B_\lambda$ 是普通背景基 $B$ 的放大或缩小。

当 $\lambda \in [0,1)$ 时,背景基 $B$ 锁定的区域 $R(B)$ 包含 $B_\lambda$ 锁定的区域 $R(B_\lambda)$,即 $R(B_\lambda) \subset R(B)$;

当 $\lambda=1$ 时,满足关系 $R(B_\lambda)=R(B)$;

当 $\lambda \in (1,\infty)$ 时,$R(B_\lambda) \supset R(B)$。

综上,对于区域 $R$ 关系满足:

$$\begin{cases} R(B_\lambda) \subset R(B), \lambda \in [0,1) \\ R(B_\lambda)=R(B), \lambda=1 \\ R(B_\lambda) \supset R(B), \lambda \in (1,\infty) \end{cases} \tag{3}$$

定义 9 设 $O$ 为第 $i$ 类背景基 $[B]_i$ 的重心,其中 $m=|[B]_i|$,$j \in m$,$b_{ij} \in [B]_i$,则 $b_{ij}^{\lambda}=O-\lambda(O-b_{ij})$。

综上,本文提出 FPCA 算法,具体描述如下。

```
算法 2:FPCA 算法
输入:测试数据 Test,参数向量 Λ = (λ_1,λ_2,…,λ_c), i = 1,2,…,
      c, λ_i > 0,各类别背景基集合 B
输出:分类结果 predict
0: predict = ∅, B_λ = ∅
1: for x in Test: #遍历测试数据中的元素
2:   for B_class in B: # B_class = IBBEA(Train_data)
3:     B_i = IBBEA(B_class)
4:     计算 B_i 的几何中心 o:
           o = (1/k) Σ_{j=1}^{k} b_j, b_j ∈ B_i
5:     var= Λ.index( B_i )
6:     for B_item in B_i :
7:         B_item = o - var × (o - B_item) #生成 λ-背景基
8:         B_λ.append(B_item)
9:     end for
10:    计算(α,β_j)= αβ_j^T,其中 α=o-x, b_j=b_j-x, j=1,2,…,|B_λ|
11:    if (α,β_j) < 0:
12:        predict.append( new_category)
13:    else:
```

```
14:        if isInnerPoint(x, B_λ ):
15:            predict. append(x. class)
16:            B_class. append(x)
17:    end for
18: end for
```

FPCA 算法基于 IBBEA 算法提取出的类别背景基集合对数据进行分类,其详细解释描述如下。

输入参数:经过数据预处理的测试数据 Test,参数向量 $\boldsymbol{\Lambda}$ ,多个类别背景基集合 $B$。

参数声明:初始化 *predict* 为空集用于存放预测结果,初始化 $B_\lambda$ 为空集用于存放 $\lambda$-背景基。

步骤 1)循环遍历测试数据 Test 中的每个元素 $x$。

步骤 2)对于每个元素 $x$,进一步遍历多类别背景基集合 $B$ 中的每个类别背景基 ***B_class***,***B_class*** 由 IBBEA 算法对训练数据进行背景基提取生成。

步骤 3)用于步骤 16)新增背景基元素后,再次使用 IBBEA 算法对新增后的背景基进行更新,得到 $i$ 类别背景基 $B_i$ 。

步骤 4)计算 $B_i$ 的几何中心 $o$,中心点 $o$ 根据背景基中所有元素的位置计算得到。

步骤 5)找到参数向量 $\boldsymbol{\Lambda}$ 中 $B_i$ 背景基的流量预测限定值 *var*。

步骤 6)~9)遍历 $B_i$ 中的每个元素 *B_item*,根据定义 9 中的公式变换生成 $\lambda$-背景基,并添加到 $B_\lambda$ 集合中。

步骤 10)~16)计算当前样本元素 $x$ 分别与几何中心 $o$ 和其他背景基元素 $b$ 之间的距离,记作$(\alpha, \beta)$,并判断$(\alpha, \beta)$的值是否小于 0,若是,则认为当前数据 $x$ 为新类别数据,将类别添加到 *predict* 集合中;否则,通过调用 isInnerPoint( ) 函数判断 $x$ 是否属于 $B_\lambda$ 背景基集合的内点,若是,则认为 $x$ 属于当前 ***B_class*** 背景基类别,将其类别添加到 *predict* 集合中,并将 $x$ 添加到 ***B_class*** 中,在步骤 3 中使用 IBBEA 算法更新新增元素后的 ***B_class*** 背景基。

步骤 17)~18)结束循环,并返回构建好的预测结果 *predict*。

设 $M$ 为样本个数,$N$ 为因素个数,$p$ 为基点个数。当样本数量较少时,$p^2$ 与 $M$ 的数量级接近;当样本数量远远大于 $N$ 和 $p^2$时,此时 $N$ 和 $p^2$ 可以忽略不计,时间复杂度为 $O(M\times N\times p^2)\approx O(M)$。因此,FPCA 算法在处理大样本数据时具有明显优势。

## 3　实验与分析

本文测试实验基于 Windows 10 操作系统实现;硬件配置:Intel(R) Core(TM) i7-6500U CPU @ 2.50 GHz、16 GB RAM;编程语言:Python3. 8。

### 3.1　实验数据集

1)NSL-KDD 数据集

NSL-KDD 数据集是对 KDDCUP99 数据集进行优化后的版本,剔除了原数据集中的部分重复和冗余记录,且训练数据和测试数据分布非常合理,有效提升了数据质量和实验效果,在比较不同的入侵检测方法时评估结果对比其他数据集更为准确,广泛应用于流量异常检测领域。数据集包括 KDDTest+, KDDTest-21、KDDTrain+和 KDDTrain+_20Percent 4 个文件,其中 KDDTest-21 和 KDDTrain+_20Percent 是 KDDTrain+和 KDDTest+的子集。该数据集的每个样本为一个连接记录,包含 41 个特征、1 个类别标签和 1 个分数。实验过程中,选用 NSL-KDD 数据集中的 KDDTrain+作为训练数据集,KDDTrain+_20Percent 作为验证数据集,KDDTest+作为测试数据集,实验所使用数据集的流量分布情况如表 1 所示。

**表 1　NSL-KDD 数据集**

**Table 1　Dataset of NSL-KDD**

| 流量类型 | 训练集/条 | 验证集/条 | 测试集/条 |
|---|---|---|---|
| 正常流量 | 67 343 | 13 449 | 9 711 |
| 攻击流量 | 58 630 | 11 743 | 12 833 |

2)CICIDS2017 数据集

通信安全机构与加拿大网络安全研究所于 2017 年在真实环境收集 1 周网络数据得到的数据集,包含良性和最新的常见攻击,其中星期 1 收集的数据仅包含正常数据,星期 2~星期 5 收集正常数据和攻击数据,旨在收集真实、先进且多样性的网络数据用于评测现有入侵检测系统的可靠性。数据集中主要包含 8 个文件,共 3 119 345 条样本,每个样本 78 个特征。其中每条样本包含 1 个正常标签和 14 个攻击标签,本文将全部攻击数据与部分正常数据(避免失衡)拼接为实验数据集,其数据分布如表 2 所示。

**表 2　CICIDS2017 数据集**

**Table 2　Dataset of CICIDS2017**

| 类别 | 描述 | 样本数/条 |
|---|---|---|
| 正常数据 | BENIGN | 529 918 |
| 攻击数据 | Dos Hulk | 231 073 |
| | PortScan | 158 930 |
| | DDoS | 128 027 |
| | DoS GoldenEye | 10 293 |
| | FTP-Patator | 7 938 |
| | SSH-Patator | 5 897 |
| | DoS Slowloris | 5 796 |
| | DoS Slowhttptest | 5 499 |
| | Bot | 1 966 |
| | Web Attack-Brute Force | 1 507 |
| | Web Attack-XSS | 652 |
| | Infiltration | 36 |
| | Web Attack-SQL Injection | 21 |
| | Heartbleed | 11 |
| 总计 | - | 1 087 564 |

### 3.2 实验评估指标

依据入侵检测标准对所提算法与相关算法进行入侵检测系统的检测性能分析,准确率(accuracy, Acc)、召回率(Recall)、精准率(Precision)、*F*1-*score* 和平均绝对误差(mean absolute error, MAE)等指标评价检测性能,相关数据根据异常检测分类混淆矩阵计算得出,一般而言,这些评价指标的值越大代表异常检测效果越好,即分类效果越好。

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

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

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

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{7}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \widehat{y_i} | \tag{8}$$

其中,TP(true positive)代表正常样本中被正确检测为正常样本的样本数;FP(false positive)代表异常样本中被错误检测为正常样本的样本数;TN(true negative)代表异常样本中被正确检测为异常样本的样本数;FN(false positive)代表正常样本中被错误检测为异常样本的样本数;$\widehat{y_i}$ 为模型预测值,$y_i$ 为真实值。

### 3.3 实验结果与分析

1)实验参数设置

KNN 离群点检测的目的是去除原始数据集中远离大部分数据的离散数据,获得一个密度高的小规模高质量数据,利用这种高质量的数据提取出更加有效精确的背景基点集合,生成对各个类别数据有更高识别度的认知包,提高异常检测的精度。其核心参数 $k$ 值表示确定一个数据点是否为离群点的最近邻居数量,通过测试不同的 $k$ 值对检测性能的影响,发现在 $k = 15$ 的时候异常检测的性能最佳,因此实验中选择 $k$ 的值为 15,去除离群点前后的数据集大小对比如表 3 所示。

**表 3 去除离群点前后的正常数据和异常数据量对比**

**Table 3 Comparison of normal and abnormal data volumebefore and after removal of outliers**

| 种类 | 去除前/条 | 去除后/条 |
|---|---|---|
| 正常数据 | 67 343 | 63 770 |
| 异常数据 | 58 630 | 56 358 |

所提方法存在多个影响分类性能的参数,贝叶斯优化(Bayesian optimization, BO)算法在迭代次数少、运行速度快以及对非凸问题表现稳健等方面具有优势。因此,本文使用 BO 算法对模型参数进行优化调整,寻找最优的参数组合,利用测试数据集对检测结果进行性能验证。参数优化具体结果如表 4 所示。

**表 4 实验参数**

**Table 4 Experimental parameter**

| 参数 | 参数说明 | 取值 |
|---|---|---|
| $\eta$ | 离群因子阈值 | 1.17 |
| $n$ | 初始背景基点个数 | 15 |
| $\lambda_{\text{normal}}$ | 正常流量预测限定值 | 1.25 |
| $\lambda_{\text{abnormal}}$ | 异常流量预测限定值 | 1.57 |

图 3 给出了在表 4 设定参数条件下不同训练次数下模型平均绝对误差。从图 3 可知,训练集和测试集在模型训练初期的 *MAE* 较高,随着训练次数的增加,*MAE* 快速下降,最终模型收敛并趋于稳定,说明模型可以在给定的数据集中完成模型收敛并形成较好的稳定输出模型。

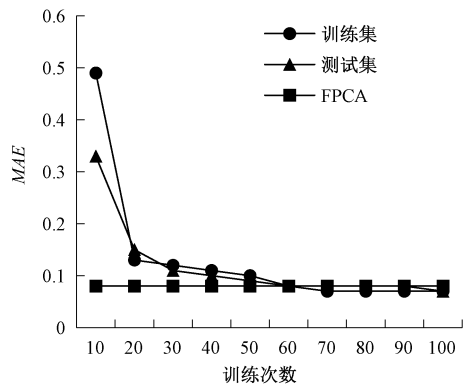


图 3 不同训练次数下的平均绝对误差对比

Fig. 3 Comparison of MAE with different training times

2)KNN 离群检测实验

为验证离群点对本文方法性能的影响,实验通过对比本文所提方法在使用 KNN 离群点检测算法去除离群点后的数据集和使用未去除离群点的数据集在检测性能上的差异,来说明离群点检测算法对提升本文方法异常检测有很大提高,结果如图 4 所示。

从图 4 中可知,使用 KNN 离群点检测算法后,本文方法对异常数据检测的各项指标都有很大提高,准确率提高了 7.81%,精确率提高了 5.78%,召回率提高了 6.41%,综合指标 *F*1-*score* 提高了 6.1%,说明去除离群点对本文方法有着极为重要的影响。背景基通过对特征因素之间的精细计算提取得到,进而生成各个类别单位认知包,离群点的存在就是由于其中某个特征因素的值偏离了大部分同类数据。离群点检测可以剔除数据集中这些远离高密度数据的异常点,减少离散数据中异常因素值在背景基提取过程中占据的权重。

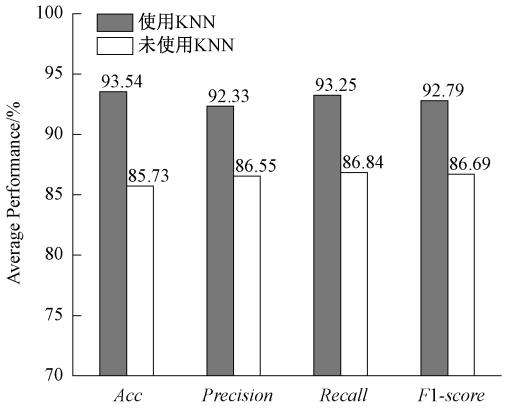


图 4　去除离群点前后的检测性能对比

Fig. 4　Comparison of detection performance before and after removal of outliers

3)扩展性实验

为验证本文方法的可扩展性,将所提方法应用于不同规模的测试集,检测其准确率、精确率、召回率和 *F*1-*score* 指标的变化趋势,以评估其是否能够在不同数据集上维持稳定的性能表现。从测试数据集中随机抽取数据,并确保每次抽取后不放回,构建 3 个规模各异的校准数据集,分别为包含 25%校准数据的测试集、包含 50%校准数据的测试集以及包含全部校准数据的测试集,每个校准数据集都分别提取了正常流量和异常流量数据对应的背景基集合,详细情况如表 5 所示。

**表 5　不同校准比例的测试数据集**

**Table 5　Test datasets with different calibration ratios**

| 流量类别 | 校准比例 | | |
|---|---|---|---|
| | 25% | 50% | 100% |
| Normal | 16 836 | 33 612 | 67 343 |
| Abnormal | 14 657 | 29 315 | 58 630 |
| Normal Background group | 163 | 347 | 697 |
| Abnormal Background group | 101 | 193 | 416 |

图 5 呈现了不同规模测试集的检测性能对比,可以观察到随着测试集规模的逐步扩大,正常流量背景基和异常流量背景基都得到了相应的增长和扩展,且在准确率、精确率、召回率和 *F*1-*score* 上均能保持稳定。这是由于背景基提取算法分别对正常流量和异常流量的因素特征进行综合计算,每新增一条流量数据都会用已有背景基对新的增量包进行验证和更新:若新增数据符合已有背景基规则,可由当前背景基表达,则过滤新增数据,不参与背景基的提取计算过程;否则,若新增数据无法由当前背景基表达得到,说明其中部分因素特征规则未被已有背景基学习,则将新增数据加入背景基的提取计算过程,学习其中新的特征规则,完善背景基与各类别流量的匹配规则。随着数据量的增加,背景基集合得到进一步的补充和完善,对异常流量的识别度始终可以保持稳定。综上所述,本文方法在不同规模的测试数据集上均表现出良好的适应性,充分证明了其具备优秀的扩展性。

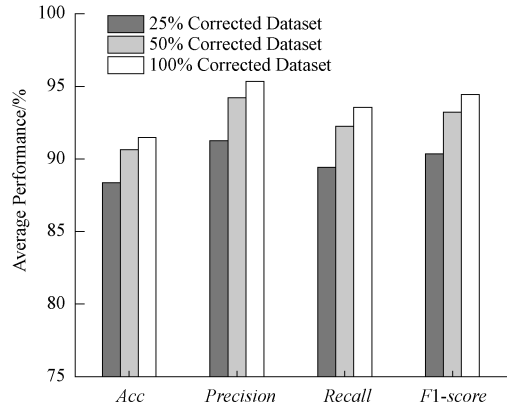


图 5　不同规模测试集检测性能对比

Fig. 5　Comparison of different test datasets detection performance

4)与经典机器学习方法对比

为验证本文方法在流量异常检测方面的性能,验证其有效性,实验将所提方法与 3 种经典机器学习方法进行综合横向性能对比,3 种机器学习方法分别是多层感知机(multilayer perceptron, MLP)、支持向量机和随机森林(random forest, RF),实验结果如表 6 所示。

**表 6　NSL-KDD 数据集上对比结果**

**Table 6　Results of comparison on the NSL-KDD dataset**

(%)

| 方法 | 准确率 | 精确率 | 召回率 | *F*1-*score* |
|---|---|---|---|---|
| MLP | 91.45 | 91.38 | 90.55 | 90.96 |
| SVM | 91.36 | 90.35 | 90.93 | 90.64 |
| RF | 92.53 | 91.89 | 91.26 | 91.57 |
| 本文方法 | 92.48 | 92.74 | 91.63 | 92.18 |

由表 6 中可知,本文方法对异常流量的检测精确率、召回率和 *F*1-*score* 均高于其他 3 种检测方法,平均增加 1.53%、0.72%和 1.12%,说明所提方法在异常检测中可以有效减少将正常流量误判为异常的情况,拥有较低的误报率和较好的综合检测性能。与 MLP 和 SVM 相比,所提方法在准确率方面优势稍弱,分别提高 1.03%和 1.12%,相比 RF 稍逊一筹,但仍超过 92%。分析上述结果可知:本文方法首先在数据预处理阶段去除各类别数据中的离群点,既抑制了数据中的噪声又降低了离群点对背景基提取过程产生的负面影响;其次,通过挖掘流量类别与因素特征之间的耦合关系进行多角度学习得到背景基,形成对正常流量和攻击流量具有较高识别度的类别认知包,可有效识别正常与异常样本的核心区别。综上所述,本文方法相较于经典机器学习方法有一定的性

能上升,特别是对异常流量样本的检测漏报情况较少,综合检测性能得到提升。

5)可行性验证

为了验证本文方法的可行性,将 CICIDS2017 数据集中星期 1 的正常数据与星期 2~星期 5 的攻击数据按照 3∶7 比例划分测试集和训练集,并与最新的深度学习算法进行比较,实验流程与 NSL-KDD 数据集一致,实验结果如表 7 所示。

**表 7 CICIDS2017 数据集上对比结果**

**Table 7 Results of comparisonon the CICIDS2017 Dataset**

(%)

| 方法 | 准确率 | 精确率 | 召回率 | *F*1-*score* |
|---|---|---|---|---|
| AE | 84.19 | 83.84 | 84.28 | 84.06 |
| DSEBM | 82.83 | 82.21 | 82.87 | 82.54 |
| DAGMM | 83.15 | 83.04 | 83.76 | 83.39 |
| 本文方法 | 87.64 | 87.16 | 88.49 | 87.82 |

由表 7 中可知,本文方法的准确率为 87.64%,精确率为 87.16%,召回率为 88.49%,*F*1-*score* 达到 87.82%,各项性能指标都优于其他模型。主要原因在于所提方法通过背景基提取来提取各类别流量因素特征之间的关系,而不是只对单个流量本身的特征进行学习,利用因素特征之间的关系进行流量异常检测,可以有效提高模型的表征能力,使模型对异常流量的识别能力得到大大提高。综上所述,所提方法在真实的网络环境中依然可以保持较高的检测率,进一步验证了其在实际应用中的可行性。

6)与现有流量异常检测模型对比

为验证本文方法在流量异常检测方面的性能,进一步证明其有效性,将所提方法与近年最新流量异常检测模型在 NSL-KDD 和 CICIDS2017 两个数据集上进行实验比较,对比结果如表 8 所示。

**表 8 与现有流量异常检测模型对比结果**

**Table 8 Results of comparison with existing traffic anomaly detection models**

(%)

| 方法 | 数据集 | 准确率 | 精确率 | 召回率 | *F*1-*score* |
|---|---|---|---|---|---|
| 文献[21] | NSL-KDD | 91.74 | 91.62 | 91.86 | 91.54 |
| | CICIDS2017 | 86.33 | 87.11 | 86.47 | 86.79 |
| 文献[22] | NSL-KDD | 91.22 | 91.75 | 91.18 | 91.46 |
| | CICIDS2017 | 87.74 | 87.39 | 87.26 | 87.32 |
| 文献[23] | NSL-KDD | 91.39 | 92.57 | 91.63 | 92.09 |
| | CICIDS2017 | 86.73 | 86.21 | 86.54 | 86.37 |
| 本文方法 | NSL-KDD | 92.48 | 92.74 | 91.63 | 92.18 |
| | CICIDS2017 | 87.64 | 87.16 | 88.49 | 87.82 |

从表 8 中对比结果可知,在 NSL-KDD 数据集上,本文方法以 92.48%的准确率、92.74%的精确率和 92.18%的 *F*1-*score* 领先其他模型,展现出最优性能。在 CICIDS2017 数据集上,尽管准确率和精确率非最高,但本文方法以 88.49%的召回率和 87.82%的 *F*1-*score* 表现最优。文献[21]利用滑动窗口和小波变换技术对网络流量数据进行小波分解和重构,保留原始流量中的特征信息,对不同特征数据具有较高的检测性能和良好的泛化能力,但由于其计算复杂度高且对参数设置敏感,导致在某些情况下性能有所波动,进而影响了 *F*1-*score*。文献[22]以重建损失和马氏距离为损失函数的自编码器完成异常流量检测,虽然无需标签数据且可解释性好,但由于对与正常样本差异较小的异常行为检测效果不佳,导致漏检和误报的情况,影响了最终的检测结果。文献[23]使用 CNN 进行特征提取并结合 TSODE 进行特征选择,这种方法增强了模型的分类效果,但由于参数调优困难以及可能的过拟合问题,导致其在某些指标上并未达到最优。相比之下,本文方法通过去除离群点并学习因素特征间的关系,有效提取了流量数据的背景基,从而形成了有效的分类基础知识,在 NSL-KDD 和 CICIDS2017 两个数据集上 *F*1-*score* 达到了最高的 92.18% 和 87.82%,充分说明该方法的有效性,能有效提高流量异常检测性能。

## 4 结 论

为了解决现有异常检测方法中存在的问题,提出一种因素空间背景基的流量异常检测基点分类方法。将因素空间理论应用到流量异常检测场景,所提方法在去除流量数据中的离群点后对因素特征间的关系进行学习,充分挖掘存在于因素特征与流量类别的关联信息,提取出可以表达同类别流量数据的背景基,形成各类别流量数据的单位认知包,并将其作为基点分类算法中识别异常流量的基础知识完成异常检测。NSL-KDD 和 CICIDS2017 数据集上实验结果表明,所提方法可以有效提高对异常流量识别能力,其综合指标高于其他同类型方法;与经典机器学习方法检测结果相近,但更适合大样本数据场景的时间复杂度需求;而由于存在少数类攻击样本不足的问题,模型对少数类攻击数据知识的学习不足,导致对少数类攻击的检测效率不高。下一步工作将对所提方法进一步优化和改进,使其在一些复杂的应用场景中展现出更好地性能。

## 参考文献

[1] 宋雅洁,步兵. 基于网络流量与设备状态的 CBTC 入侵检测系统[J]. 中国安全科学学报, 2019(S2): 161-167.
SONG Y J, BU B. An intrusion detection system based on network traffic and device states for CBTC[J]. China

Safety Science Journal, 2019(S2):161-167.

[2] 陈雪倩,步兵. 基于网络流量和数据包的 CBTC 入侵检测系统[J]. 中国安全科学学报,2019(S2):154-160.
CHEN X Q, BU B. An intrusion detection system for CBTC based on network traffic and packets[J]. China Safety Science Journal, 2019(S2):154-160.

[3] 张维,步兵,王洪伟. 基于 KF 的列控系统数据篡改攻击检测方法[J]. 中国安全科学学报, 2019(S1):32-37.
ZHANG W, BU B, WANG H W. An intrusion detection system for CBTC based on network traffic and packets[J]. China Safety Science Journal, 2019(S1):32-37.

[4] 王琳琳,刘敬浩,付晓梅. 基于极限学习机与改进 K-means 算法的入侵检测方法[J]. 计算机工程与科学, 2018,40(8):1398-1404.
WANG L L, LIU J H, FU X M. Intrusion detection method based on extreme learning machine and improved K-means algorithm [J]. Computer Engineering & Science, 2018, 40(8):1398-1404.

[5] 陈万志,徐东升,张静. 工业控制网络入侵检测的 BP 神经网络优化方法[J]. 辽宁工程技术大学学报(自然科学版), 2019,38(1):82-87.
CHEN W ZH, XU D SH, ZHANG J. BP neural network optimization method for industrial control network intrusion detection[J]. Journal of Liaoning Technical University(Natural Science), 2019, 38(1):82-87.

[6] 陈万志,李东哲. 结合白名单过滤和神经网络的工业控制网络入侵检测方法[J]. 计算机应用, 2018, 38(2):363-369.
CHEN W ZH, LI D ZH. Intrusion detection method in industrial control network combining white list filtering and neural network [J]. Computer Engineering and Science, 2018, 38(2):363-369.

[7] 王华忠,杨智慧,颜秉勇,等. 融合 PCA 和 PSO-SVM 方法在工控入侵检测中的应用[J]. 科技通报, 2017, 33(1):80-85.
WANG H ZH, YANG ZH H, YAN B Y, et al. Application of fusion PCA and PSO-SVM method in industrial control intrusion detection [J]. Bulletin of Science and Technology, 2017, 33(1):80-85.

[8] 梁辰,李成海,周来恩. PCA-BP 神经网络入侵检测方法[J]. 空军工程大学学报(自然科学版), 2016, 17(6):93-98.
LIANG CH, LI CH H, ZHOU L EN. A PCA-BP neural network-based intrusion detection method[J]. Journal of Air Force Engineering University (Natural Science Edition), 2016, 17(6):93-98.

[9] YAN W Z, YU L J. On accurate and reliable anomaly detection for gas turbine combustors: a deep learning approach[J]. arXiv preprint arViv:1908.09238, 2019.

[10] FIORE U, PALMIERI F, CASTIGLIONE F, et al. Network anomaly detection with the restricted Boltzmann machine[J]. Neurocomputing, 2013, 122:13-23.

[11] SEOK S, KIM H. Visualized malware classification based-on convolutional neural network[J]. Journal of the Korea Institute of Information Security, 2016, 26(1):197-208.

[12] 吕佩吾,葛雅川,李楠,等. 基于卷积神经网络的工控协议 Modbus TCP 异常检测[J]. 信息安全研究, 2019, 5(7):635-638.
LYU P W, GE Y CH, LI N, et al. Abnormal detection in Modbus TCP based on convolutional neural network[J]. Journal of Information Security Research, 2019, 5(7):635-638.

[13] 李熠,李永忠. 基于自编码器和极限学习机的工业控制网络入侵检测算法[J]. 南京理工大学学报(自然科学版), 2019, 43(4):408-413.
LI Y, LI Y ZH. Intrusion detection algorithm for industrial control networks based on auto encoder and extreme learning machine [J]. Journal of Nanjing University of Science and Technology, 2019, 43(4):408-413.

[14] 王竹晓, 张彭彭,李为,等. 基于深度 Q 网络的电力工控网络异常检测系统[J]. 计算机与现代化, 2019(12):114-118.
WANG ZH X, ZHANG P P, LI W, et al. Electric power industrial control network anomaly detection system based on deep Q Network [J]. Computer and Modernization, 2019(12):114-118.

[15] 梁欣怡,行鸿彦,侯天浩. 基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法[J]. 电子测量与仪器学报, 2022, 36(10): 65-73.
LIANG X Y, XING H Y, HOU T H. CNN-BiLSTM network intrusion detection method based oneself-supervised feature enhancement [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 65-73.

[16] 罗佳,黄晋英. 生成式对抗网络研究综述[J]. 仪器仪表学报,2019,40(3):74-84.
LUO J, HUANG J Y. Review of generative adversarial networks [J]. Chinese Journal of Scientific Instrument, 2019, 40(3):74-84.

[17] 汪培庄. 因素空间与数据科学[J]. 辽宁工程技术大学学报(自然科学版),2015,34(2):273-280.
WANG P ZH. Factor space and data science [J]. Journal of Liaoning Technical University (Natural

Science), 2015, 34(2):273-280.
[18] 吕金辉,刘海涛,郭芳芳,等. 因素空间背景基的信息压缩算法[J]. 模糊系统与数学,2017,31(6):82-86.
LYU J H, LIU H T, GUO F F, et al. Information compression algorithm based on factor space background [J]. Fuzzy Systems and Mathematics, 2017, 31(6):82-86.
[19] 蒲凌杰,曾繁慧,汪培庄. 因素空间理论下基点分类算法研究[J]. 智能系统学报,2020,15(3):528-536.
PU L J, ZENG F H, WANG P ZH. Base point classification algorithm based on factor space theory[J]. CAAI Transactions on Intelligent Systems, 2020, 15(3): 528-536.
[20] 陈万志,张国满,王天元. 基于特征耦合泛化的流量异常检测方法[J]. 电子测量与仪器学报,2024,38(2):120-130.
CHEN W ZH, ZHANG G M, WANG T Y. Traffic anomaly detection method based on feature coupling generalization[J]. Journal of Electronic Measurement and Instrumentation,2024,38(2):120-130.
[21] DUAN X Y, FU Y, WANG K. Network traffic anomaly detection method based on multi-scale residual classifier[J]. Computer Communications, 2023, 198:206-216.
[22] YANG D H, HWANG H. Unsupervised and ensemble-based anomaly detection method for network security[C]. Proceedings of 14th International Conference on Knowledge and Smart Technology. Piscataway: IEEE Press, 2022:75-79.
[23] FATANI A, ABD E M, DAHOU A, et al. IoT intrusion detection system using deep learning and enhanced transient search optimization[J]. IEEE Access, 2021, 9:123448-123464.

## 作者简介

**陈万志**(通信作者),2015 年于辽宁工程技术大学(中国测绘科学研究院联合培养)获得博士学位,现为辽宁工程技术大学副教授,硕士生导师,主要研究方向为人工智能与智能信息处理、网络与信息安全和工控软件与数据分析。
E-mail: chenwanzhi@ lntu. edu. cn

**Chen Wanzhi** (Corresponding author) received his Ph. D. degree from Liaoning Technical University (China Academy of Surveying and Mapping Science Joint Cultivation) in 2015, respectively. Now he is an associate professor and master's supervisor in Liaoning Technical University. His main research interests include artificial intelligence and intelligent information processing, network and information security and industrial control software and data analytics.

**任鹏江**,2021 年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为网络安全和入侵检测。
E-mail: 2878402446@ qq. com

**Ren Pengjiang** received his B. Sc. degree from Liaoning Technical University in 2021. Now he is a M. Sc. candidate at Liaoning Technical University. His main research interests include network security and intrusion detection.