智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

耦合演化采样和深度解码的可解释网络流量异常检测模型

孙俊, 谢振平, 王洪波

引用本文:
孙俊,谢振平,王洪波. 耦合演化采样和深度解码的可解释网络流量异常检测模型[J]. 智能系统学报, 2023, 18(5): 1070–1078.
SUN Jun, XIE Zhenping, WANG Hongbo. An explainable network traffic anomaly detection model with coupled evolutionary sampling and deep decoding[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(5): 1070–1078.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202211035

DOI: 10.11992/tis.202211035
网络出版地址: https://kns.cnki.net/kcms2/detail/23.1538.TP.20230615.1214.004.html

耦合演化采样和深度解码的可解释网络流量异常检测模型

孙俊[1,2], 谢振平[1,2], 王洪波[3]

（1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122; 2. 江南大学 江苏省媒体设计与软件技术重点实验室, 江苏 无锡 214122; 3. 拓尔思天行网安信息技术有限责任公司, 北京 100089）

摘　要: 针对现有网络流量异常检测模型缺乏可解释性的问题, 本研究提出了耦合演化采样和深度解码的可解释网络流量异常检测模型。首先, 引入演化采样学习抽取代表特征样本, 依此实现了强可解释性的样本编码过程; 其次, 构建了可解释的演化采样样本编码过程和不可解释的深度神经网络解码过程的耦合学习模型; 最后, 使用样本编码结果和重构误差进行异常检测。在 NSL-KDD 和 CICIDS2017 数据集上与现有方法的实验比较结果表明, 该方法可显著提升模型可解释性和模型规模效率, 并能取得与现有最优方法同等水平的检测性能。此外, 上述新的学习策略, 也可为可解释机器学习方法研究提供一种极具特色的技术方案参考。

关键词: 机器学习; 无监督学习; 流量异常检测; 深度神经网络; 可解释性; 演化采样; 深度编码; 自编码器
中图分类号: TP391　　**文献标志码:** A　　**文章编号:** 1673−4785(2023)05−1070−09

中文引用格式: 孙俊, 谢振平, 王洪波. 耦合演化采样和深度解码的可解释网络流量异常检测模型 [J]. 智能系统学报, 2023, 18(5): 1070–1078.
英文引用格式: SUN Jun, XIE Zhenping, WANG Hongbo. An explainable network traffic anomaly detection model with coupled evolutionary sampling and deep decoding[J]. CAAI transactions on intelligent systems, 2023, 18(5): 1070–1078.

An explainable network traffic anomaly detection model with coupled evolutionary sampling and deep decoding

SUN Jun[1,2], XIE Zhenping[1,2], WANG Hongbo[3]

(1. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; 2. Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China; 3. TRS Topwalk Information Techololgy Co., Ltd, Beijing 100089, China)

Abstract: Regarding the lack of explainability in existing network traffic anomaly detection models, this study proposed an explainable network traffic anomaly detection model with coupled evolutionary sampling and deep decoding. First, evolutionary sampling learning is introduced to extract representative feature samples, whereby a strongly explainable sample encoding process is implemented. Second, a coupled learning model of the explainable evolutionary sample encoding process and the unexplainable deep neural network decoding process is constructed. Finally, anomaly detection is performed using the sample encoding results and reconstruction errors. The experimental analysis on NSL-KDD and CICIDS2017 datasets are executed for our model and some existing methods, and corresponding results show that our model can significantly improve model explainability and scale efficiency and achieve the same level of detection performance as existing optimal methods. In addition, our proposed joint learning strategy may provide a highly distinctive scheme reference for the development of explainable machine learning methods.

Keywords: machine learning; unsupervised learning; traffic anomaly detection; deep neural network; explainability; evolutionary sampling; deep enconding; autoencoder

互联网的广泛应用和飞速发展推动了社会快速发展, 同时也带来一系列网络安全问题。网络流量作为网络信息传输和交互的载体, 包含了大

收稿日期: 2022−11−21.　**网络出版日期:** 2023−06−15.
基金项目: 国家自然科学基金项目(62272201, 61872166).
通信作者: 谢振平. E-mail: xiezp@jiangnan.edu.cn.

量网络行为相关信息[1]。其中, 异常网络流量是指会影响网络正常运行的流量, 主要有两类[2]: 一是由网络结构不合理和网络使用不当造成的异常; 二是由 DDos 或 SQL 注入等网络攻击造成的异常。若能及时发现并捕获异常网络流量, 就能够更好地保障网络的安全运行。网络流量异常检测通过将各种异常检测方法用于网络流量数据分析, 并在此基础上发现异常网络流量并产生报警。

传统网络流量异常检测包括基于分类[3]、统计[4-5]、聚类和信息论[6] 这 4 大类[7] 方法。这些方法使许多机器学习算法能够应用于网络流量异常检测。但随着网络流量数据规模的变大, 机器学习算法已经无法满足现实需求。随着近些年深度学习的快速发展, 基于重构[8-9] 和对抗[10-11] 等的无监督模型在网络流量异常检测领域取得了优异的结果, 其学习正常样本的潜在特征, 可解决带标签数据难以获取的问题[12]。除此之外, 还有一些无监督学习方法, 如深度玻尔兹曼机和深度信念网络 (deep belief network, DBN)[13] 也都被广泛地应用。

自编码器[4,14](autoencoder, AE) 拥有优秀的数据重构和特征表征能力。因此, 许多学者围绕基于自编码器的算法进行研究, 并提出了许多行之有效的模型。其中 Zong 等[4] 提出了深度自编码高斯混合模型 (deep autoencoder Gaussian mixture model, DAGMM), 采用 AE 和高斯混和模型 (Gaussian mixture model, GMM) 来进行网络流量异常检测。Zhai 等[15] 提出了深度结构能量模型 (deep structured energy based model, DSEBM), 采用深度能量结构对数据进行分布建模, 将集成的不同类型的数据与 AE 连接, 从而降低信息损失。Gong 等[16] 提出了深度记忆自编码器模型 (memory-augmented deep autoencoder, MemAE), 采用了 Memory 模块来扩充 AE。此外, 生成对抗网络[17](generative adversarial network, GAN) 最初作为图像生成领域的模型取得了很大的成功。由于其出色的性能被越来越多地运用于网络流量异常检测领域。其中, Schlegl 等[10] 提出了基于 GAN 的异常检测模型, 是 GAN 用于异常检测的开山之作。黄训华等[18] 提出了多模态对抗学习异常检测 (multimodal GAN, MMGAN), 将对抗学习扩充到多个模态上。Audibert 等[19] 提出了无监督异常检测 (unsupervised anomaly detection, USAD), 将 GAN 来优化 AE 的训练。

同时, 越来越多的学者关注到网络流量异常检测的可解释问题, 并围绕这个问题提出了许多可解释增强的异常检测模型。其中, Ting 等[20] 提出了孤立分布核 (isolation distributional kernel, IDK), IDK 本质上是一个特征核, 它可以将离群点很好地表征出来, 从而进行可解释的异常检测。Chen 等[21] 提出了插值高斯描述子 (interpolated Gaussian descriptor, IGD), 采用了插值高斯描述子的方法来训练一类高斯异常分类器, 高斯异常分类器用来引导样本的重构, 并依此增强重构误差的可解释性。

现有的基于深度学习的网络流量异常检测模型大多侧重于构建样本重构前后相似或相异关系, 忽略了可解释的特征表征; 而现有的可解释网络流量异常检测模型可分为构造浅层可解释模型和在深度学习模型中加入可解释模块这两种思路, 这两类思路均忽略了可解释模块和深度学习的耦合关系。上述问题在很大程度上限制了网络流量异常检测的实际应用, 因此, 本文提出了耦合演化采样[22] 和深度解码的可解释网络流量异常检测模型 (an explainable network traffic anomaly detection model with coupled evolutionary sampling and deep decoding, CESDDM)。演化采样获取代表性特征样本, 本文将其称为编码基, 实现了强可解释的样本编码, 且将可解释的样本编码与不可解释的深度解码过程耦合学习, 然后样本编码结果和重构误差进行异常判定。本文的主要贡献包括:

1) 引入演化采样样本编码替换原始编码结构, 以获得强可解释性的编码基。

2) 实现了可解释的演化采样样本编码过程与不可解释的深度解码过程的耦合学习。

3) 将样本编码结果直接参与网络流量异常判定, 以此获得强可解释性的判定结果。

1　新模型方法

1.1　网络结构

本文提出的 CESDDM 由深度编码和演化采样两个模块构成。其中深度编码由样本编码和深度解码两部分构成。本文网络结构如图 1 所示。在训练阶段, 首先使用样本编码替换原始编码过程并进行特征提取, 之后引入演化采样和深度解码的耦合学习策略, 深度解码学习正常流量样本的模式以最小化重构误差, 演化采样学习更新编码基以获取最优代表性网络流量样本。在测试阶段, 给定测试流量样本, CESDDM 仅使用编码基中记录的有限数量的正常流量进行样本编码和深度解码, 最后基于样本编码结果和重构误差与阈值的比较来进行异常检测。CESDDM 将可解释性的演化采样和不可解释的深度解码耦合构造, 从而构建了一个可解释的深度模型。

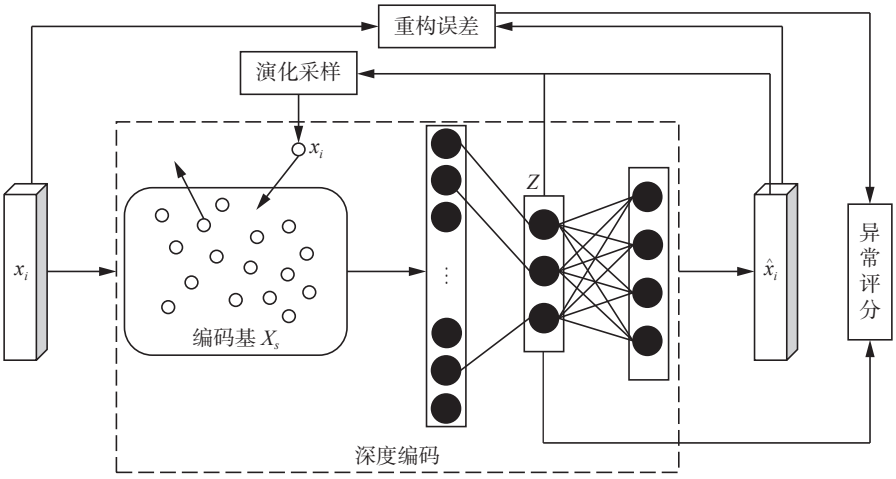


图 1　本文网络结构

Fig. 1　Our model framework

1.2　样本编码

在 CESDDM 中，样本编码可以被视为一个相似度选择器，即将原始流量样本与编码基中最高的t个相似度作为特征进行提取。样本编码的结果将作为深度解码的输入和异常判别的标准。

给定一组原始流量样本$X=\{x_i|i=1,2,\cdots,N_O\}$和一组编码基$X_S=\{x_j|j=1,2,\cdots,N_S\}$，则对于任意原始流量样本$x_i$，其样本编码过程如下：

$$z=\text{top}(\{T=k(x_i,x_j;\theta_1)|j=1,2,\ \cdots,N_S\};t) \quad (1)$$

式中：$k(\cdot)$是高斯核函数；θ_1是高斯核函数的超参数；top(·)是一个选择函数，用于选择T中最大的t个值。

传统自编码器的编码结构是线性的，而样本编码是非线性的。这使得样本编码存在一定的随机性，但这种随机性不会导致编码结果的随机化，反而是样本编码的结果稳定的关键。具体而言，编码基X_S是一个能够代表原始流量样本分布和特征的子集，而样本编码是一个提取相似度特征的过程，那么对于正常流量样本而言，必然能够在编码基中找到与之相似的样本，所以这种特征提取是稳定的。也正是利用样本编码这一特点，其能够直接参与网络流量异常判别。同时为了尽可能地保留样本编码的线性结构，top(·)得到的编码结果并非按照相似度大小排列，而是需要对齐X_S中的顺序。选择的数量t对模型的影响会在 2.4.4 中详细说明。

1.3　演化采样和深度解码的耦合训练

演化采样学习 (evolutionary sampling learning, ESL) 是一种通用的机器学习框架，旨在从原始流量样本中采样得到一组称为编码基的代表样本，编码基可以用作概率分布的建模。其适用于在一定概率框架内转化为密度估计的机器学习问题。受 ESL 的启发，CESDDM 利用 ESL 变形方法来耦合深度解码的学习。

在一定概率框架内，对于一组原始流量样本X，必然能够找到一组编码基X_S，其能够代表X的内在特征，包含了关于X的近似最优信息。最优化编码基是 ESL 的目标，为了训练得到最优的编码基，预定义任意原始流量样本x在原始分布和编码基上的密度估计：

$$f(x,\theta_1;X)=\frac{1}{N_o}\sum_{i=1}^{N_O}k(x,x_i;\theta_1) \quad (2)$$

$$p(x,\theta_1;X_S)=\frac{1}{N_S}\sum_{j=1}^{N_S}k(x,x_j;\theta_1) \quad (3)$$

式中：N_O表示原始流量样本X的样本量；N_S表示编码基X_S的样本量；$k(\cdot)$是一个核函数，通常被考虑为高斯核函数。$k(\cdot)$中距离选用余弦距离代替常用的欧氏距离，余弦距离的定义如下：

$$D(A,B)=1-\frac{A\cdot B}{\|A\|_2\,\|B\|_2} \quad (4)$$

给定原始流量样本$X=\{x_i|i=1,2,\cdots,N_o\}$，将$X$中前$N_S$个样本作为初始编码基。将$x_i$作为样本编码的输入，提取其相似度特征$z$作为深度解码的输入。深度解码将$z$重构，并最小化重构误差，其结构如下：

$$\hat{x}_i=\sigma(Wz+b) \quad (5)$$

式中：W是深度解码的权重；b是深度解码的偏置；$\sigma(\cdot)$是 ReLU 激活函数。深度解码在完成对原始流量样本重构的同时保留z的梯度，并由此得到Δ_z，Δ_z的定义如下：

$$\Delta_z=\left(\tilde{z}-z\right)\cdot l \quad (6)$$

其中，$\tilde{z}$是z经过深度解码反向传播之后的值，l是权重常数。利用Δ_z来更新式 (2)：

$$f(x_j,\theta_1;X)=a\cdot f(x_j,\theta_1;X)\pm(1-a)\cdot\omega\cdot S(\Delta_z) \quad (7)$$

其中：x_j表示z对应X_S中的样本，a是权重系数，$S(\cdot)$

是 Sigmoid 激活函数, ω是对$S(\cdot)$的放缩值。

由于神经网络反向传播的性质, 本文针对Δ_z的取值分情况讨论。具体而言, 当Δ_z为正值时, 认为其对应编码基中的x_j是好的代表样本, 应该增加其权重, 使其不易被替换; 相反地, 当Δ_z为负值时, 认为其对应编码基中的x_j是差的代表样本, 应该减小其权重, 使其更易被替换。演化采样学习正是通过上述方法抽取代表性特征样本。

未被样本编码选中的x_j的更新公式如下:

$$f(x_j,\theta_1;X)=a\cdot f(x_j,\theta_1;X) \tag{8}$$

之后从编码基X_S中选择候选更新样本x_r:

$$r=\max\{p(x_j,\theta_1;X_S)-f(x_j,\theta_1;X)\} \tag{9}$$

最后, 根据是否满足下式判断是否将x_r替换为x_i:

$$R\sim[0,1]\leqslant\left\{0,\frac{f(x_r,\theta_1;X)\cdot p(x_i,\theta_1;X_S)}{f(x_i,\theta_1;X)\cdot p(x_r,\theta_1;X_S)}\right\} \tag{10}$$

在 CESDDM 中, 演化采样和深度解码的学习互相引导耦合构造, 其关系如图 2 所示。具体而言, 演化采样更新编码基的过程依赖于深度解码的反向传播, 这是深度解码引导演化采样的过程; 深度解码的输入由样本编码直接提供, 这是演化采样引导深度解码的过程。这种互相引导学习的结构使得可解释的演化采样和不可解释的深度解码可以耦合构造。

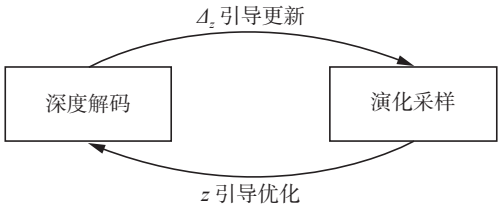


图 2　演化采样和深度解码的耦合训练过程

Fig. 2　Coupled training process of evolutionary sampling and deep decoding

深度解码使用均方误差 (MSE) 作为损失函数, 具体定义如下:

$$M_l=\frac{1}{n}\sum_{n=1}^{n}(x_i-\hat{x}_i)^2 \tag{11}$$

ESL 的训练目标以及X_S对X的近似精度计算如下:

$$p(x,\theta_1;X_S)\approx f(x,\theta_1;X) \tag{12}$$

$$E_l=\frac{1}{N_S}\sum_{x_j\in X_S}\frac{f(x_j,\theta_1;X)}{p(x_j,\theta_1;X_S)} \tag{13}$$

1.4　异常判别和可解释性分析

在测试阶段, 给定测试样本x_t, 本文的异常判定如下:

$$y_t=\begin{cases}1,\ \text{mean}(z_t)\leqslant\mu_1\ ,\ R(x_t,\hat{x}_t)\geqslant\mu_2\\0,\ \text{其他}\end{cases} \tag{14}$$

式中: $y_t=1$表示测试样本判定为异常, $y_t=0$表示测试样本判定为正常, μ_1和μ_2为预设的阈值, 同时满足两部分条件的测试样本会被判定为异常。$\text{mean}(z_t)$是样本编码结果的均值, $R(\cdot)$是测试样本的重构误差, 具体定义如下:

$$R(x_t,\hat{x}_t)=\|x_t-\hat{x}_t\|^2 \tag{15}$$

在训练阶段, CESDDM 与其他网络流量异常检测模型均使用编码结构进行特征提取, 与之不同的是, CESDDM 使用演化采样样本编码过程替换了原始的深度神经网络编码过程。原始的编码结构虽然有着很强的特征表征能力, 但缺乏可解释性。相比而言, 演化采样样本编码提取原始流量样本与编码基的相似度特征进行编码, 使得本文的编码过程是具备强可解释性的。

在测试阶段, 目前大多数网络流量异常检测算法仅使用重构误差作为异常判定。而 CESDDM 在使用重构误差的同时引入 mean(·)用作异常判别。具体而言, 正如上文提到的, 编码基是一组原始网络流量样本的代表性样本, 其包含了正常网络流量的近似最优信息。因此, 正常测试样本总能在编码基中找到与其相似度较高的样本; 相反地, 异常测试样本由于其离群、孤立的特性, 很难在编码基中找到与其相似度较高的样本。所以, 样本编码的结果作为可解释部分直接参与异常判定。

2　实验结果与分析

2.1　实验数据集与超参数

实验选取了网络流量异常检测领域两个具有代表性的公开数据集, 一个为经典的网络流量数据集 NSL-KDD 数据集; 另一个为数据量较大, 且各种异常类型比较全面的 CICIDS2017 数据集。本文使用了两个数据集的各 20000 条正常样本作为训练集, 实验测试集的具体构成如表 1 所示。

表 1　实验测试集情况

Table 1　Experimental datasets information

数据集	正常样本	异常样本	异常比例
NSL-KDD	9711	12833	0.57
CICIDS2017	22334	22334	0.50

NSL-KDD 数据集[23]: 著名的网络流量数据集 KDD99 的改进版本, 其解决了原数据中存在大量冗余的问题, 是网络流量异常检测通用的一个经典数据集。本文采用正常流量样本作为训练集, 每个样本包括 41 种特征, 其中 34 种连续特征, 7 种分类特征 (离散型数据)。实验将字符型特征转化为数值型特征, 然后利用方差选择法选

择得到 12 维特征。

CICIDS2017 数据集[24]：加拿大网络安全研究所于 2017 年采集并公开的网络流量数据集，其中包括了正常流量与常见攻击导致的异常流量。异常流量包括暴力文件传输协议 (FTP)、暴力安全外壳协议 (SSH)、拒绝服务 (DDoS) 等。本文使用正常流量样本作为训练集，每个样本包含 78 种特征，将包含脏数据的特征剔除，然后利用方差选择法选择得到 57 维特征。

经过基于不同参数的比较实验，CESDDM 在 NSL-KDD 和 CICIDS2017 数据集下的基本运行超参数设定如表 2 所示。其中，LR是 CESDDM 的学习率，θ_1是高斯核函数的超参数，N_S 是编码基的样本量，t是样本经过样本编码后的维度。

表 2　模型在 2 个数据集上的超参数设置

Table 2　Hyperparameter setting of models on two datasets

数据集	LR	θ_1	N_S	t
NSL-KDD	0.001 0	0.08	200	20
CICIDS2017	0.000 5	0.06	200	20

2.2　评价指标

与本领域相关成果一样，本文采用精确率、召回率和 F_1-score 等评价指标，由异常检测混淆矩阵得到相关数据，如表 3 所示。通常，我们期望这些评价指标的值尽可能大。

表 3　异常检测分类混淆矩阵

Table 3　Confusionmatrixforanomalydetectionclassification

真实类别	检测异常	检测正常
异常样本	T_P	F_N
正常样本	F_P	T_N

精确率：体现了检测结果为异常样本中异常样本的比例，计算方法如下：

$$P = \frac{T_P}{T_P + F_P} \tag{16}$$

召回率：体现了异常样本被正确识别的比例，计算方法如下：

$$R = \frac{T_P}{T_P + F_N} \tag{17}$$

F_1-score：基于精确率和召回率两项指标计算，其作用在于当精确率和召回率都无法比较模型的综合性能时 (例如：召回率高，但精确率低)，F_1-score 作为精确率与召回率的一种折中方式来比较模型的综合性能。计算方法如下：

$$F_1 = \frac{1}{a/P + (1+a)/R} \tag{18}$$

式中：a可以实现精确率和召回率的折中，一般情况下，a取值为 0.5。

2.3　对比方法

1) OC-SVM[25](one-class support vector machine)：一种经典的基于核函数的异常检测模型，通过学习正常样本和异常样本之间的边界来进行异常检测。OC-SVM 对于小数据集能取得较好的效果，泛化能力较强，但对于数据量较大，维度较高的数据集却很难取得满意的结果。

2) IF[12](isolation forests)：一种经典的识别离群数据的异常检测模型，它将异常点定义为“容易被孤立的离群点”，即那些分布稀疏且距离密度高的集合较远的点。在数据空间内，若一个区域内只有离群点，则表示数据点落在此区域的概率较低，因此可以判定落在此区域的点是异常点。即，IF 的理论基础有两点：正常样本数量远大于异常样本的数量；异常样本的特征值与正常样本的差异很大。若不满足条件，则 IF 对于该类数据的识别效果较差。

3) AE[26]：一种经典的基于深度学习的异常检测方法，通过编码器将数据压缩，然后通过解码器将其重构，最后基于重构误差进行异常检测。

4) DSEBM[15]：一种基于深度学习的异常检测方法，其在不同的网络层之间积累能量，并依此判断数据是否异常。DSEBM 充分利用了训练过程中的信息来检测异常，但需要正常数据与异常数据具有较大的差异。

5) DAGMM[4]：一种基于自编码器的异常检测模型，由基于 AE 的数据网络和基于 GMM 的密度估计网络组成。前者通过 AE 训练得到数据的低维表示，后者对低维表示进行密度估计，最终模型通过比较估计的样本能量和预先设定的阈值来进行异常检测。DAGMM 将低维表示作为 GMM 的输入来弥补数据压缩中的信息损失，但模型需要高质量的训练集来进行训练。

6) MemAE[16]：一种基于自编码器的异常检测模型，由 AE 和 Memory 两个模块构成，前者通过编码器得到数据的低维表示，后者查找 Memory 中与低维表示最相关的内存项，并将其作为解码器的输入，最后通过比较重构误差与预先设定的阈值来进行异常检测。MemAE 通过 Memory 模块来扩大异常样本的重构误差，但模型存在训练不充分的问题。

2.4　实验结果与分析

2.4.1　收敛性分析

CESDDM 引入了耦合演化采样和深度编码的学习策略，使用样本编码替换传统的编码结构。对于新的模型结构，需要验证其收敛性。本

小节将通过基于 NSL-KDD 数据集的实验结果从理论和实验两个方面验证模型的收敛性。图 3 和图 4 分别是演化采样和深度解码的收敛情况。

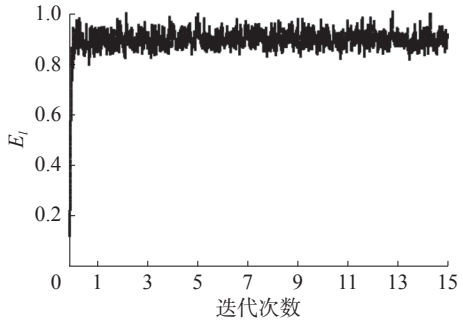


图 3　ESL 收敛散点图

Fig. 3　Convergence scatter of ESL

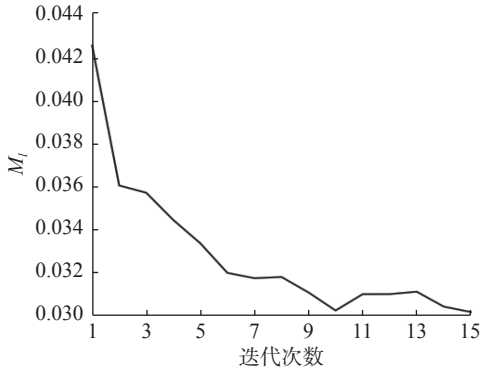


图 4　MSE 收敛曲线

Fig. 4　MSE convergence curves

演化采样的主要目的是获取原始概率分布的最优近似, 即最优化编码基。基于演化计算和抽样的概念, 演化采样可以在有限的样本数量和给定的采样条件下获取概率最优的编码基。通过最优化编码基与原始概率分布之间的相似度, 编码基可以稳定渐近地 (在概率上) 收敛于原始概率分布。根据图 3 中E_l的变化散点图所示, ESL 明显可以很快地收敛到一个稳定的区间内, 但在这个区间内E_l会存在一定的震荡。综上所述, ESL 能够很好地收敛, 训练得到的编码基可以很好地表征原始流量样本。

深度解码的主要目的是学习正常流量样本的特征, 即最小化重构误差。原始的自编码器, 通过深度神经网络进行特征提取, 并最小化重构误差。其理论基础是: 自编码器仅学习正常样本的特征信息, 即正常样本能够被很好地重构, 而异常样本难以被重构。CESDDM 使用演化采样样本编码过程替换掉原始的神经网络编码过程。样本编码学习正常样本的相似度特征, 正常样本的相似度特征远大于异常样本, 所以深度解码根据这类特征可以很好地重构正常样本, 而异常样本则难以被重构。因此, 样本编码并未破坏原始自编码器的理论基础, 理论上 CESDDM 仍能收敛。根据图 4 中深度解码M_l收敛曲线所示, 深度解码能够收敛, 这从实验角度论证了上述观点, 即样本编码未破坏自编码器的理论基础, 其结构仍具有收敛性。

综上所述, CESDDM 的两个目标函数均能在理论和实验中得到收敛, 这充分表明了 CESDDM 的逻辑可行性。

2.4.2　异常判别的可解释分析

本小节将从实验的角度论述 CESDDM 的可解释性。CESDDM 通过引入耦合演化采样和深度解码的学习策略来构建一个可解释模型。其中样本编码是 CESDDM 具有可解释性的关键, 实验通过正常流量样本和异常流量样本不同的样本编码结果来论述 CESDDM 的可解释性。

在 NSL-KDD 上的实验结果如图 5 所示, 其中图 5(a) 表示正常流量样本编码结果, 图 5(b) 表示异常流量样本编码结果。从图中可以看出, 正常流量样本的样本编码结果远大于异常流量样本的样本编码结果。这说明了正常流量样本总能从编码基中找出与其相似度较高的样本, 而异常流量样本由于其离群、孤立的性质而无法找出与其相似度较高的样本。CESDDM 利用这一特点作为异常判别的一部分, 并使得这种判别方式具备了较强的可解释性。

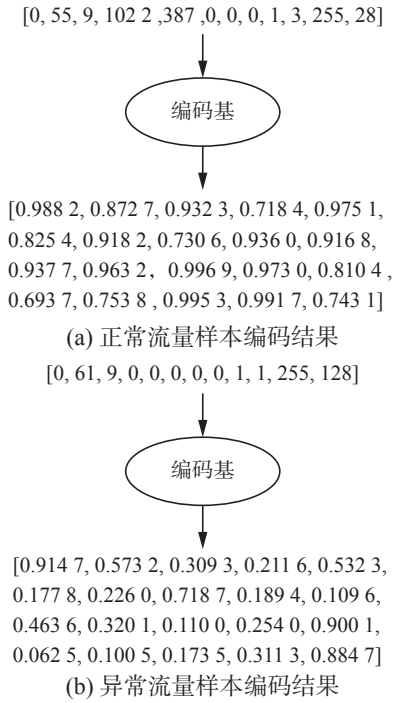


图 5　样本编码示例

Fig. 5　Sample encoding example

2.4.3　对比实验

NSL-KDD: 实验结果如表 4 所示。从表中可以看出, CESDDM 在 NSL-KDD 数据集上的精确率为 84.49%、召回率为 96.64%, F_1-score 为 90.16%。1) 对比机器学习算法 (OC-SVM 和 IF) 在 NSL-KDD 数据集上, 机器学习在得到很好训练的情况下, 由于受到数据集维度和维度的影响, IF 的精确率和召回率分别比 CESDDM 低 4.25% 和 7.40%, OC-SVM 的精确率和召回率分别比 CESDDM 低 10.70% 和 18.48%。2) 对比深度学习算法 (AE、DSEBM、DAGMM 和 MemAE), CESDDM 的精确率和 F_1-score 比其中效果最好的 MemAE 低 6.43% 和 1.30%, 但召回率比 MemAE 高 4.64%。CESDDM 的参数量远低于其他深度学习模型, 这是由于样本编码的浅层结构相较于其他深度学习算法的深层编码结构, 网络节点的参数要少很多。具体而言, 在 NSL-KDD 数据集上, 深度学习算法的效果要优于传统机器学习算法, CESDDM 在考虑了模型可解释性的情况下仍能保持较好的效果。这说明, 耦合演化采样和深度编码的学习策略可以在保证模型性能的情况下, 构造一个可解释的模型, 同时可以大幅较少模型的参数量。

表 4　NSL-KDD 数据集异常检测结果

Table 4　Anomaly detection results on NSL-KDD dataset

模型	精确率/%	召回率/%	F_1-score/%	参数量
OC-SVM	73.79	78.16	75.19	—
IF	80.24	89.24	84.50	—
AE	79.28	86.84	82.89	18 215
DSEBM	82.87	68.21	74.83	161 138
DAGMM	80.01	91.01	85.15	18 277
MemAE	**90.92**	92.00	**91.46**	18 977
CESDDM	84.49	**96.64**	90.16	**4 145**

CICIDS2017: 实验结果如表 5 所示。从表中可以看出, CESDDM 在 CICIDS2017 数据集上的精确率、召回率和 F_1-score 分别为 63.86%、79.14% 和 70.69%。1) 对比传统的机器学习方法 (OC-SVM 和 IF), 其中效果较好的 IF 的各项指标均低于 CESDDM。2) 对比深度学习算法 (AE、DSEBM、DAGMM 和 MemAE), 可以看出其中效果最好的 MemAE 也难以得到优秀的性能。这表明在复杂的数据集中, 网络流量异常检测仍然十分困难。CESDDM 在保持与深度学习算法同等水平的情况下, 构造了可解释的模型并且在参数量上具有一定优势。

表 5　CICIDS2017 数据集异常检测结果

Table 5　Anomaly detection results on CICIDS2017 dataset

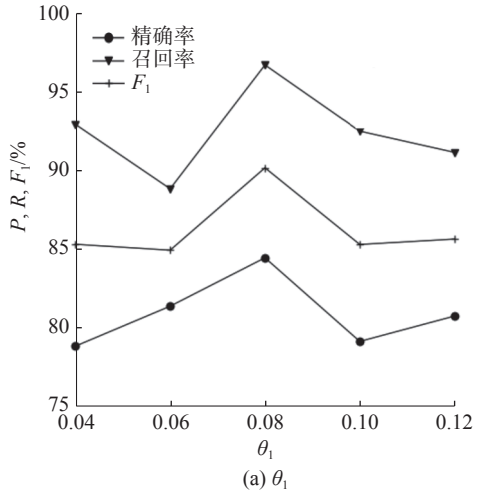
模型	精确率/%	召回率/%	F_1-score/%	参数量
OC-SVM	66.21	60.67	63.32	—
IF	63.12	71.39	66.99	—
AE	63.96	74.89	68.99	13 859
DSEBM	66.48	68.95	67.69	151 886
DAGMM	**66.50**	72.99	69.50	13 921
MemAE	63.55	**79.75**	**70.73**	14 921
CESDDM	63.86	79.14	70.69	**13 797**

综上所述, CESDDM 在不同的数据集下, 通过耦合演化采样和深度编码的学习策略, 使得本文模型具备较高的可解释性, 同时保证了与现有模型同等性能并减少了模型的参数量。从而很好地验证了 CESDDM 的有效性和特色优势。

2.4.4　参数敏感性实验

本节通过对参数θ_1、t和N_S的不同取值进行实验以论证其对模型的影响。经过相应的测试, 实验将θ_1的取值范围设定为 [0.04,0.12], t的取值范围 [10,30], N_S的取值范围设定为 [100,300], 在 NSL-KDD 数据集上进行实验, 具体的实验结果如下:

θ_1: 实验结果如图 6(a) 所示。从图中可以看出, 在θ_1不同的取值情况下, 精确率会在 [78.92%, 84.49%] 之间波动, 波动范围在 5.57% 以内; 召回率在 [88.84%. 96.64%] 之间波动, 波动范围在 7.80% 以内, 而 F_1-score 由于精确率和召回率的波动而波动。这种波动是由于θ_1的取值会直接影响两个样本相似度的计算, 从而影响正常流量样本和异常流量样本的分割。所以 CESDDM 对于θ_1的设定是敏感的。



(a) θ_1

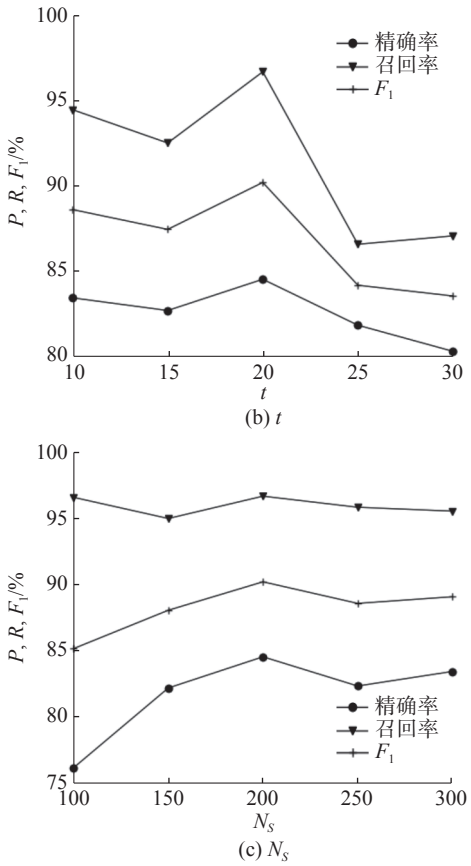


图 6　参数敏感性实验

Fig. 6　Sensitivity experiment of parameters

t: 实验结果如图 6(b) 所示。从图中可以看出, 在t不同的取值情况下, 精确率会在 [80.27%, 84.49%] 之间波动, 召回率会在 [86.54%, 96.64%] 之间波动, F_1-score 随着精确率和召回率的波动而波动。当t值在 [10,20] 之间时, 波动范围较小, 而t为 25 和 30 时精确率和召回率会有明显地下降。这是因为当t增大到一定维度后, 原始流量样本的样本编码和重构会变得不准确和困难。所以 CESDDM 对t的取值有一定阈值, 超过阈值的t会影响模型的性能。

N_S: 实验结果如图 6(c) 所示。从图中可以看出, 在N_S不同的取值情况下, 精确率会在 [76.10%, 84.49%] 之间波动, 召回率会在 [94.95%,96.64%] 之间波动。当N_S为 100 时, 精确率较低, 其余情况下精确率和召回率波动很小。这是由于, 当N_S为 100 时编码基无法很好地表征原始流量样本, 从而导致精确率降低。当N_S在合适的区间内, 编码基能够很好地表征原始流量样本。所以 CESDDM 对于N_S的设定是不敏感的。

3　结束语

本文提出了一种耦合演化采样和深度解码的可解释网络流量异常检测模型 (CESDDM)。本文学术价值在于, 使用演化采样样本编码替换原始的编码结构, 且实现了可解释样本编码和不可解释的深度解码的耦合学习, 以增强模型的可解释能力, 上述模型为可解释机器学习研究提供了一个较为特色新颖的技术思路。在本领域两个典型的公开数据集上的对比实验结果表明, CESDDM 可以在保持与现有最优深度学习算法同等性能的情况下, 保证模型的可解释性并减少模型的参数量。下阶段我们将考虑更为优化的耦合训练方法以实现高可解释性的同时进一步提升模型性能。

参考文献:

[1] LIU Hongyu, LANG Bo. Machine learning and deep learning methods for intrusion detection systems: a survey[J]. Applied sciences, 2019, 9(20): 4396.

[2] BHATTACHARYYA D K, KALITA J K. Network anomaly detection: a machine learning perspective[M]. Boca Raton: Crc Press, 2013.

[3] 杨月麟, 毕宗泽. 基于深度学习的网络流量异常检测 [J]. 计算机科学, 2021, 48(S2): 540–546.
YANG Yuelin, BI Zongze. Network anomaly detection based on deep learning[J]. Computer science, 2021, 48 (S2): 540–546.

[4] ZONG Bo, SONG Qi, MIN M R, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection[C]//International Conference on Learning Representations. Vancouver: OpenReview, 2018: 1−19.

[5] 席亮, 王瑞东, 樊好义, 等. 基于样本关联感知的无监督深度异常检测模型 [J]. 计算机学报, 2021, 44(11): 2317–2331.
XI Liang, WANG Ruidong, FAN Haoyi, et al. Sample-correlation-aware unsupervised deep anomaly detection model[J]. Chinese journal of computers, 2021, 44(11): 2317–2331.

[6] TAN Zhiyuan, JAMDAGNI A, HE Xiangjian, et al. A system for denial-of-service attack detection based on multivariate correlation analysis[J]. IEEE transactions on parallel and distributed systems, 2014, 25(2): 447–456.

[7] AHMED M, NASER M A, HU Jiankun. A survey of network anomaly detection techniques[J]. Journal of network and computer applications, 2016, 60: 19–31.

[8] TRAN C P, TRAN D K. Anomaly detection in POSTFIX mail log using principal component analysis[C]//2018 10th International Conference on Knowledge and Systems Engineering. Ho Chi Minh City: IEEE, 2018: 107−112.

[9] 李贝贝, 彭力, 戴菲菲. 结合马氏距离与自编码器的网络流量异常检测方法 [J]. 计算机工程, 2022, 48(4): 133–142.
LI Beibei, PENG Li, DAI Feifei. Abnormal network traffic detection method combining mahalanobis distance and autoencoder[J]. Computer engineering, 2022, 48(4): 133–142.

[10] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al.

Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]//International Conference on Information Processing in Medical Imaging. Cham: Springer, 2017: 146−157.

[11] ZENATI H, ROMAIN M, FOO C S, et al. Adversarially learned anomaly detection[C]//2018 IEEE International Conference on Data Mining. Singapore. IEEE, 2018: 727−736.

[12] ZHANG Kunzhong, KANG Xudong, LI Shutao. Isolation forest for anomaly detection in hyperspectral images[C]//2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama: IEEE, 2019: 437−440.

[13] SINGH K, MATHAI K J. Performance comparison of intrusion detection system between deep belief network (DBN)algorithm and state preserving extreme learning machine (SPELM) algorithm[C]//2019 IEEE International Conference on Electrical, Computer and Communication Technologies. Coimbatore: IEEE, 2019: 1−7.

[14] 王倩倩, 苗夺谦, 张远健. 深度自编码与自更新稀疏组合的异常事件检测算法 [J]. 智能系统学报, 2020, 15(6): 1197–1203.
WANG Qianqian, MIAO Duoqian, ZHANG Yuanjian. Abnormal event detection method based on deep auto-encoder and self-updating sparse combination[J]. CAAI transactions on intelligent systems, 2020, 15(6): 1197–1203.

[15] ZHAI Shuangfei, CHENG Yu, LU Weining, et al. Deep structured energy based models for anomaly detection[C]//International conference on machine learning. New York: PMLR, 2016: 1100−1109.

[16] GONG Dong, LIU Lingqiao, LE V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 1705−1714.

[17] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. New York: ACM, 2014: 2672−2680.

[18] 黄训华, 张凤斌, 樊好义, 等. 基于多模态对抗学习的无监督时间序列异常检测 [J]. 计算机研究与发展, 2021, 58(8): 1655–1667.
HUANG Xunhua, ZHANG Fengbin, FAN Haoyi, et al. Multimodal adversarial learning based unsupervised time series anomaly detection[J]. Journal of computer research and development, 2021, 58(8): 1655–1667.

[19] AUDIBERT J, MICHIARDI P, GUYARD F, et al. USAD: UnSupervised anomaly detection on multivariate time series[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 3395−3404.

[20] TING Kaiming, XU Bicun, WASHIO T, et al. Isolation distributional kernel: a new tool for kernel based anomaly detection[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 198−206.

[21] CHEN Yuanhong, TIAN Yu, PANG Guansong, et al. Deep one-class classification via interpolated Gaussian descriptor[J]. Proceedings of the AAAI conference on artificial intelligence, 2022, 36(1): 383–392.

[22] XIE Zhenping, SUN Jun, PALADE V, et al. Evolutionary sampling: a novel way of machine learning within a probabilistic framework[J]. Information sciences, 2015, 299: 262–282.

[23] TAVALLAEE M, BAGHERI E, LU Wei, 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.

[24] SHARAFALDIN I, HABIBI L A, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy. Portugal: SCITEPRESS-Science and Technology Publications, 2018: 108−116.

[25] LI Kunlun, HUANG Houkuan, TIAN Shengfeng, et al. Improving one-class SVM for anomaly detection[C]//Proceedings of the 2003 International Conference on Machine Learning and Cybernetics. Xi'an: IEEE, 2004: 3077−3081.

[26] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special lecture on IE, 2015, 2(1): 1–18.

作者简介：

孙俊，硕士研究生，主要研究方向为网络流量异常检测、机器学习。

谢振平，教授，博士生导师，主要研究方向为知识计算与认知学习。获教育部科技进步一等奖、全国商业科技进步特等奖/一等奖等科研奖励。主持或主要参与完成国家、省部级科研项目 9 项，获授权发明专利 6 项。发表学术论文 50 余篇。

王洪波，高级工程师，主要研究方向为网络安全软件及系统。