基于深度加权特征学习的网络安全态势评估

杨宏宇[1,2], 张梓锌[2], 张　良[3]

[1] 中国民航大学安全科学与工程学院 天津 中国 300300
[2] 中国民航大学计算机科学与技术学院 天津 中国 300300
[3] 亚利桑那大学信息学院 图森 美国 AZ 85721

摘要 计算机网络高速发展的同时也带来了许多的安全问题, 对网络安全进行有效的网络安全态势评估对于掌握网络整体的状态并帮助管理人员全面掌握整体态势具有重要意义。然而, 现有的网络安全态势评估方法存在特征要素提取困难、准确率低、时效性差的问题。针对这些问题, 提出一种面向网络威胁检测的基于深度加权特征学习的网络安全态势评估方法。首先, 考虑到单个稀疏自动编码器进行特征提取时无法很好的拟合不同攻击的分布, 从而影响威胁检测准确率的缺点, 构建一个基于并行稀疏自动编码器的特征提取器提取网络流量中的关键信息, 并将其与数据原始特征进行融合。其次, 为了更多的关注网络流量中的关键信息, 采用注意力机制改进双向门控循环单元网络, 对网络中的威胁进行检测并统计每种攻击类型的发生次数以及误报消减矩阵。然后, 根据误报消减矩阵修正每种攻击类型的发生次数, 并结合威胁严重因子计算得到威胁严重度。最后, 根据威胁严重度和每种攻击类型的威胁影响度确定网络安全态势值以获取网络安全态势。本文选取 NSL-KDD 数据集进行实验验证, 实验结果显示本文方法在测试集上达到了 82.13%的最高准确率, 召回率、*F*1 值分别达到了 83.36%、82.74%。此外, 通过消融实验进一步验证了所提出的并行稀疏自动编码器提取特征和注意力机制加权特征两种改进方法的有效性。与经典态势评估方法 SVM、LSTM、BiGRU、AEDNN 等的对比实验也证明, 该方法能够高效、全面地评估网络安全的整体态势。

关键词 并行稀疏自动编码器; 注意力机制; 威胁严重因子; 误报消减矩阵; 网络安全态势评估
中图法分类号 TP309 DOI 号 10.19363/J.cnki.cn10-1380/tn.2022.07.03

Network Security Situation Assessment Based on Deep Weighted Feature Learning

YANG Hongyu[1,2], ZHANG Zixin[2], ZHANG Liang[3]

[1] Department of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China
[2] Department of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China
[3] Department of Information, University of Arizona, Tucson AZ 85721, USA

Abstract The rapid development of computer network also brings many security problems, network security situation assessment is of great significance for mastering the overall state of the network and helping managers fully grasp the overall situation. However, the available network security situation assessment methods have difficulties in extracting feature elements, low precision and the poor timelines. To tackle this problem, a network security situation assessment method based on deep weighted feature learning for network threat detection was proposed. Firstly, considering the disadvantage of a single sparse automatic encoder to fit the distribution of different attacks when extracting features, which affects the accuracy of threat detection, a feature extractor based on a parallel sparse auto-encoder was built to extract key data of network traffic and integrate them with the original features. Then, to pay more attention to the key information in the network traffic, the attention mechanism was used to improve the improved Bi-directional Gate Recurrent Unit. The network threat was tested by the testing set and the occurrence number of each attack type and the false alarm reduction matrix were counted. Then, the occurrence number of each attack type was corrected according to the false alarm reduction matrix, and the threat severity was calculated by combining the threat severity factor of each attack type. Finally, the network security situation was determined according to the threat severity and the threat impact level of each attack type. On the data sets of the NSL-KDD, the experimental results show that the proposed method achieves the highest precision of 82.13% in the test dataset, and the recall and *F*1 scores reach 83.36%, and 82.74% respectively. The ablation experiment further verifies the effectiveness of the proposed two improved methods: parallel sparse automatic encoder to extract features and attention mechanism weighted features. Besides, the comparative experiment with the classical situation assessment methods such as SVM, LSTM, BiGRU, AEDNN also prove that the proposed method can assess the whole situation of network security efficiently and comprehensively.

通讯作者: 杨宏宇, 博士, 教授, Email:yhyxlx@hotmail.com。
本课题得到国家自然科学基金民航联合研究基金资助项目(No.U1833107)资助。
收稿日期: 2021-05-19; 修改日期: 2021-08-08; 定稿日期: 2022-05-11

Key words parallel sparse auto-encoder; attention mechanism; threat severity factor; false alarm reduction matrix; network security situation assessment

1 引言

随着通信技术和云计算技术的发展, 现今几乎所有的行业都开始应用计算机网络进行办公[1]。与此同时, 恶意攻击或破坏造成的网络安全事件也越来越普遍, 网络和信息系统面临着众多网络攻击的威胁[2]。因此, 全面掌握网络的整体安全状态是一个亟待解决的热点问题。网络安全态势评估(network security situation assessment, NSSA)可以根据相关安全事件构建合适的模型, 进而评估网络系统整体所遭受的威胁程度, 帮助安全管理人员掌握当前网络状况[3-4]。

目前, 国内外相关研究已取得一定成果[5]。Lu 等[6]将网络安全态势分为主机安全态势和网络攻击态势两部分, 设计权重和计算规则以计算网络安全态势。Agrawal 等[7]基于模糊分析网络过程评估标准的权重, 并通过模糊对称技术评估软件的安全性。此外, 还有层次分析法(analytic hierarchy process, AHP)[8-9]、集对分析法[10]、模糊数学[11]等方法, 但此类运用数学模型的方法受主观因素影响较大, 没有客观的标准。Alali 等[12]提出利用模糊逻辑推理系统改进网络安全风险评估模型, 并综合分析了脆弱性、威胁、可能性和影响等四个方面从而得出风险评估结果。杨宏宇等[13]基于自修正系数修匀法, 通过熵关联度、自适应解和时变加权马尔可夫链改进网络安全态势的预测结果。此外, 还有运用概率和知识推理的方法如贝叶斯网络[14-15]、模糊推理[16]、D-S 证据理论[17]等, 这些方法依赖于专家知识库和大量的规则推理, 在海量数据的网络环境下存在模型构建困难、操作复杂等问题。杨宏宇等[18]基于无监督学习, 提出一种通过解析多源网络流量评估网络威胁的态势评估方法。该方法具有较强的网络威胁特征识别能力, 对网络威胁态势评估有效性的提升提供了可行的思路。Hong 等[19]则将灰色关联分析理论和支持向量机(support vector machine, SVM)算法用于网络安全态势预测, 实验结果表明该模型具有更高的网络风险预测精度。但此类运用模式分类的方法在实时环境中提取特征困难, 建模时间长, 不易于理解。

为了应对日益复杂的网络威胁和攻击, 网络安全技术不断地被更新和发展, 研究人员开始尝试利用深度学习的方法研究网络安全问题。Lin 等[20]基于门控循环单元(gate recurrent unit, GRU)、双向门控循环单元(bi-directional gate recurrent unit, BiGRU)等多种神经网络模型对 UNSW-NB15 数据集进行检测, 结果表明, 与其他模型相比, BiGRU 的准确率最高。文献[21]将改进的 LSTM 应用于 KDD99 数据集, 实验证明该方法可有效地理解和评估网络安全态势。文献[22]设计了一种基于对抗学习的态势评估模型 AEDNN, 解决了传统方法面对大量数据时效率低的问题。Chakravarthi 等[23]提出一种基于深度自动编码器(auto-encoder, AE)提取特征的入侵检测方法, 得到了表征能力更强的特征, 但使用该方法训练网络模型时存在梯度消失的问题。Moradi 等[24]将基于堆叠式自动编码器提取特征的特征学习和孤立森林相结合, 获得了良好的检测结果, 但文中仅检测有无攻击发生, 无法满足攻击类型进行细分与检测需要。文献[25]将 MapReduce 和 SVM 相结合并应用于网络安全态势预测, 解决了 SVM 训练时间长的缺点, 但未对网络态势进行全面的评估, 无法反映网络的整体态势情况。Shone 等[26]将非对称深度自编码器的无监督特征学习应用于入侵检测并取得了较好的检测结果, 但该方法在少数攻击类别上的检测率为 0, 存在着攻击类型样本数失衡导致的弱检测问题。

近年来, 一些研究人员尝试用注意力机制对深度学习网络进行改进, 以提高安全检测的性能。Liu 等[27]采用基于注意力机制的深度神经网络进行 web 攻击的实时检测, 在真实的网络流量上证明了该方法的可行性。Arnav 等[28]将一种基于注意力机制的自动编码器应用于异常检测, 实验证明该方法相对于其他自动编码器变体具有更高的检测性能。Yang 等[29]用注意力机制改进 LSTM 并将其用于威胁检测, 取得了较好的检测效果。

针对目前网络安全态势评估方法在获取先验知识、提取特征、构建模型、实时性等方面存在的不足, 为了有效、全面地评估网络安全态势, 本文提出一种基于深度加权特征学习的网络安全态势评估方法。通过并行稀疏自动编码器(parallel sparse auto-encoder, PSAE)高效、准确地提取不同攻击类型的特征并与数据原始特征融合, 采用注意力机制改进 BiGRU 网络(attention-based BiGRU, ATBiGRU), 再使用改进后的网络模型(parallel sparse auto- encoder-attention-based BiGRU, PSAE-ATBiGRU)进行网络威胁检测, 根据测试结果计算网络安全态势量化值。

2 基于 PSAE 的特征提取与融合

2.1 稀疏自动编码器

自动编码器(AE)是一种无监督的特征提取算法, 其结构如图 1 所示。AE 结合了编码器以及解码器, 并使用反向传播将它们联系在一起。编码器将输入转换为低维抽象来提取原始特征并学习数据表示, 解码器接收低维表示并重建原始特征。

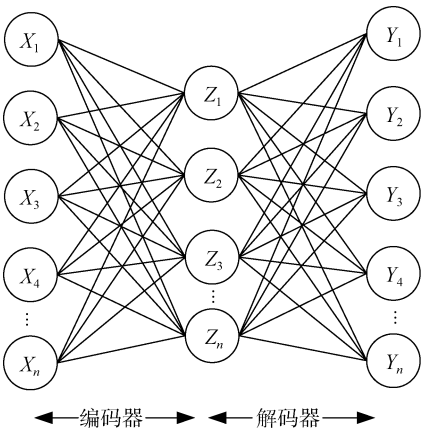

图 1 自动编码器网络结构图

Figure 1 AE's network structure

稀疏自动编码器(sparse auto-encoder, SAE)[30]是在 AE 基础上的改进。SAE 为了避免简单地从输出到输入的映射, 在隐藏层上添加了稀疏性约束, 增加模型的泛化能力, 获得更好的特征描述。SAE 通过反向传播获得权重矩阵, 选择 *Sigmoid* 函数 $g(z)=1/(1+e^{-z})$用于激活神经网络层中的神经元。神经元的稀疏性由神经元的输出决定。如果神经元的输出接近 1, 认为它是活动的。如果神经元的输出接近 0 时, 认为它是不活动的。在使用反向传播的 SAE 中, 损失函数为

$$A_{sparse}(m,x_i,y_i)=\frac{1}{2m}\sum_{i=1}^{m}\left\|x_i-y_i\right\|^2+\beta\sum_{j=1}^{K}KL\left(\rho\left\|\hat{\rho}_j\right.\right) \quad (1)$$

其中, m 指输入神经元数, K 指隐藏神经元数, x_i 指输入数据, y_i 指输出数据。在上式中, β 控制神经元的稀疏程度, ρ 表示网络中神经元的期望激活水平, $\hat{\rho}_j$ 表示第 j 个神经元的平均激活水平。此外, *KL* 散度的计算公式为

$$KL\left(\rho\,\|\,\hat{\rho}_j\right)=\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \quad (2)$$

除了稀疏约束之外, 通常还会通过 L2 正则化避免模型过拟合的问题, 因此最终的损失函数为

$$J_{sparse}=A_{sparse}(n,x_i,y_i)+\frac{\lambda}{2}\left(\sum_{k,n}\boldsymbol{W}^2+\sum_{k,n}\boldsymbol{V}^2+\sum_{k}\boldsymbol{b}_1{}^2+\sum_{k}\boldsymbol{b}_2{}^2\right) \quad (3)$$

其中, λ 指正则化参数, n 指层数, k 指当前层数, $\boldsymbol{W}$ 和 $\boldsymbol{V}$ 指权重矩阵, $\boldsymbol{b}_1$ 和 $\boldsymbol{b}_2$ 指偏置项。

2.2 PSAE 特征提取器的设计

特征学习是一种仅对属性子集的数据行为进行建模的技术, 它可有效显示检测性能与数据模型质量之间的相关性。通过使用新特征对网络进行训练, 可以提高网络分类效率和分类准确性。因此可通过特征提取与融合来增强原始特征的表征能力, 从而提高分类的准确性。

此外, NSL-KDD 数据集[31]包含多种攻击类型, 且这些类型的信息分布各不相同, 通过单个 SAE 进行特征提取时间长且无法很好的拟合不同攻击的分布。因此可用多个特征提取器分别学习每种攻击的分布规律, 更好的表达不同攻击类型之间的信息差异。

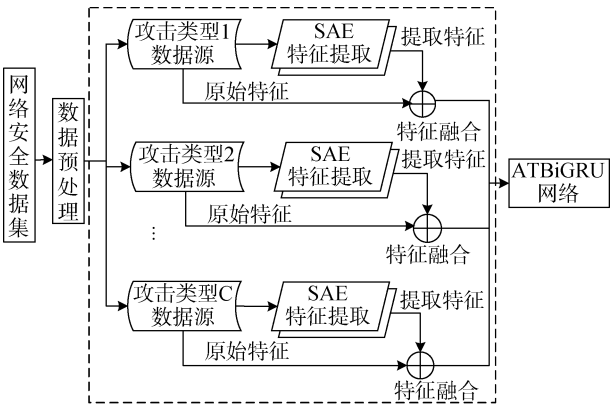

图 2 基于 PSAE 的特征提取器

Figure 2 Feature extractor based on PSAE

本文设计的基于 PSAE 的特征提取器结构如图 2 所示。首先, 将数据预处理之后的数据集按照不同的攻击类型输入 SAE 特征提取器进行特征提取。其中, SAE 隐藏层神经元的数量等于其编码器所学习的输入数据压缩表示的个数。编码器对原始数据进行压缩, 解码器重构原始输入数据的特征表示。训练完成后, 将编码器输出结果作为代表原始数据的特征, 即可完成特征提取功能。最后, 将提取的特征与原始特征融合, 输入至 ATBiGRU 模型进行训练。其中, PSAE 的训练及特征提取过程如算法 1 所示。

算法 1. PSAE 的训练及特征提取

输入: 不同网络威胁的攻击类型数据: X_0, X_1, X_2, …, X_{C-1}, 其中 X_i 表示攻击类型为 i 的所有样本数据: $X_i=\{x_{i0}, x_{i1}, \ldots, x_{i(n-1)}\}$

输出: 经过训练的 SAE 集合: S_0, S_1, S_2, …, S_{C-1}, 提取的特征: L_0, L_1, L_2, …, L_{C-1}

BEGIN

按网络威胁的攻击类型构建 SAE: S_0, S_1, S_2, …, S_{C-1}

FOR (i = 0; i< C; i ++):

WHILE (training S_i)

FOR (j = 0; j < n; j ++):

将 x_{ij} 送入 S_i 的编码器编码为 h_{ij}

将 h_{ij} 送入 S_i 的解码器解码为 y_{ij}

loss = sparse_loss (x_{ij}, y_{ij})

WHILE (提取 X_i 的特征)

FOR (j = 0; j < n; j ++):

将 x_{ij} 送入 S_i 的编码器编码为 h_{ij}

将 h_{ij} 加入到 L_i 列表中

L_i= {h_{i0}, h_{i1}, …, $h_{i(n-1)}$}

END

3 ATBiGRU 网络模型

3.1 BiGRU 网络和注意力机制

BiGRU 是 GRU 的改进版本, 其结构图如图 3 所示。BiGRU 在每个时刻的输入会经过两个方向相反的 GRU, 其输出结果综合考虑这两个 GRU 的输出。因此, BiGRU 可以学习过去和将来状态与当前状态之间的时序关系, 有助于提取更深层次的特征信息[32]。

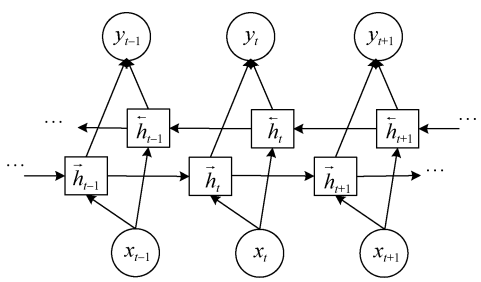

图 3 BiGRU 结构图

Figure 3 BiGRU's structure

注意力模型是 Treisman 和 Gelade 提出的类似于人脑的资源分配模型[33], 它通过对目标数据进行加权运算来突出关键特征, 较好地提升了模型的拟合效果。因此, 本文引入注意力机制, 帮助模型可以更有效地学习潜在层特征, 并对显著影响最终检测结果的关键特征进行加权, 使获得的特征信息更合理、更准确, 进而提高模型的检测精度及模型的鲁棒性。

3.2 ATBiGRU 网络设计

首先, 由于网络威胁流量属于时间序列事件, 即当前时间的攻击类型由当前时刻的数据和先前时刻的数据共同决定, 因此通过 BiGRU 可有效学习网络威胁流量间的表征关系, 增强检测网络的特征学习能力。其次, 文献[34]基于 GRU 设计了一种分层注意力网络, 该网络在选取句子关键词汇的任务上取得了较为不错的成绩。考虑到数据样本中不同时刻的特征信息冗余且对当前攻击类型的分类与检测有不同的贡献, 这与关键词汇的选取问题有着相似性, 因此采用注意力机制对关键特征加权, 实现对 BiGRU 网络模型的改进。图 4 展示了本文设计的 ATBiGRU 模型结构, ATBiGRU 的具体步骤设计如下:

步骤 1 给定若干条具有 n 个维度的样本, 其中第 i 条表示为 $\boldsymbol{X}_i$={x_{i0}, x_{i1}, …, $x_{i(n-1)}$}, 对应真实标签为 $\boldsymbol{Y}_i$。将其输入 BiGRU 网络模型, 学习样本间的时序关系, 并进行编码。通过 BiGRU 函数对前向和反向两个隐藏状态加权求和, 获得各个隐藏层的状态 $\boldsymbol{h}_{ij}$

$$\boldsymbol{h}_{ij} = \mathrm{BiGRU}\left(x_{ij}, \vec{\boldsymbol{h}}_{n-1}, \overleftarrow{\boldsymbol{h}}_{n-1}\right) \tag{4}$$

其中, $\vec{\boldsymbol{h}}_{n-1}$ 指前向隐藏层状态, $\overleftarrow{\boldsymbol{h}}_{n-1}$ 指反向隐藏层状态。

步骤 2 使用注意力机制计算每个特征应分配的概率权重, 突出网络威胁流量特征中的关键信息, 计算局部特征向量, 由公式(5)~(7)计算注意力层的权重系数以及局部特征向量

$$\boldsymbol{d}_{ij} = \tanh\left(\boldsymbol{A}_u \boldsymbol{h}_{ij} + \boldsymbol{f}_u\right) \tag{5}$$

$$a_{ij} = \frac{exp\left(d_{ij}^{\mathrm{T}} \boldsymbol{d}_u\right)}{\sum_j exp\left(d_{ij}^{\mathrm{T}} \boldsymbol{d}_u\right)} \tag{6}$$

$$\boldsymbol{s}_i = \sum_j a_{ij} \boldsymbol{h}_{ij} \tag{7}$$

其中, $\boldsymbol{d}_{ij}$ 指使用 *softmax* 函数归一化操作得到的隐藏层状态, $\boldsymbol{h}_{ij}$ 指 BiGRU 模型的输出, $\boldsymbol{A}_u$ 指加权系数, $\boldsymbol{f}_u$ 指偏置项, $\boldsymbol{d}_u$ 指随机初始化的注意力矩阵。a_{ij} 指不同概率权重和每个隐藏层状态的乘积之和, $\boldsymbol{s}_i$ 指由 $\boldsymbol{h}_{ij}$ 与 a_{ij} 加权求和得到的局部特征向量。

步骤 3 将步骤 2 的局部特征向量 $\boldsymbol{s}_i$ 输入 BiGRU 网络模型, 与步骤 2 相似, 全局特征向量 $\boldsymbol{v}$ 由概率权重 $\boldsymbol{a}_i$ 进一步计算得到, 由公式(8)~(11)计算注意力层的权重系数以及全局特征向量

$$\boldsymbol{h}_i = \mathrm{BiGRU}\left(\boldsymbol{s}_i, \vec{\boldsymbol{h}}_{n-1}, \overleftarrow{\boldsymbol{h}}_{n-1}\right) \tag{8}$$

$$\boldsymbol{d}_i = \tanh\left(\boldsymbol{A}_u \boldsymbol{h}_i + \boldsymbol{f}_w\right) \tag{9}$$

$$\boldsymbol{a}_i = \frac{exp\left(\boldsymbol{d}_i^{\mathrm{T}} \boldsymbol{d}_w\right)}{\sum_i exp\left(\boldsymbol{d}_i^{\mathrm{T}} \boldsymbol{d}_w\right)} \tag{10}$$

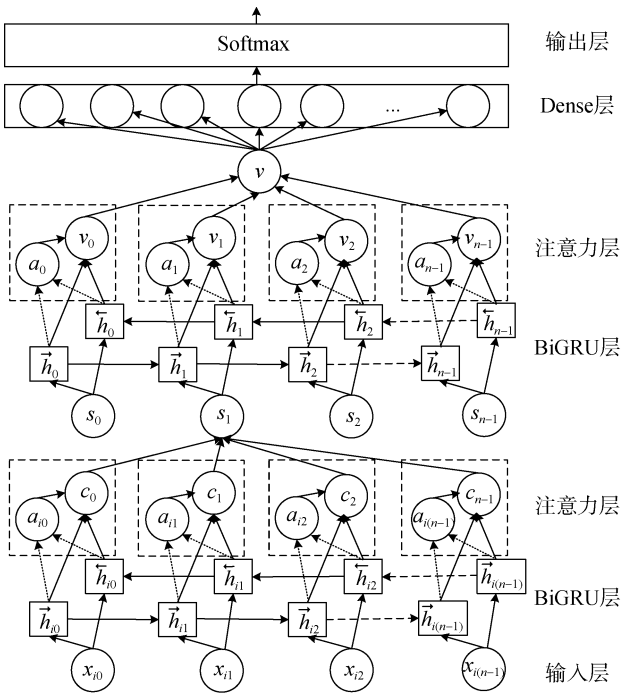

图 4　ATBiGRU 模型结构

Figure 4　ATBiGRU's structure

$$\boldsymbol{v}=\sum_{i}\boldsymbol{a}_i\boldsymbol{h}_i \tag{11}$$

其中, $\boldsymbol{A}_u$、$\boldsymbol{f}_w$ 和 $\boldsymbol{d}_w$ 分别表示第 2 层注意力机制的权重系数矩阵、偏置项和随机初始化的注意力矩阵。

步骤 4　将步骤 3 的结果通过 Dense 层进一步提取特征, 最后在 softmax 输出层输出分类结果 $Y(\boldsymbol{X}_i)$

$$Y(\boldsymbol{X}_i)=softmax(\boldsymbol{W}_w\boldsymbol{v}+\boldsymbol{b}_v) \tag{12}$$

其中, $\boldsymbol{W}_w$ 指分类器权重系数矩阵, $\boldsymbol{b}_v$ 表示分类器偏置, 输出 $Y(\boldsymbol{X}_i)$表示模型预测结果。

步骤 5　将预测结果与原始标签对比并计算误差 *loss*

$$loss=-\sum_{i=1}^{C}\boldsymbol{Y}_i\log(Y(\boldsymbol{X}_i)) \tag{13}$$

4　基于 PSAE-ATBiGRU 的网络安全态势评估方法

4.1　网络安全态势评估框架

本文提出的网络安全态势评估模型结构如图 5 所示。该模型主要包括数据预处理、PSAE-ATBiGRU 网络威胁检测和网络安全态势评估 3 个部分。

(1) 数据预处理: 对采集的网络流量数据进行特征数值化、特征约简、特征最大最小值归一化、平衡数据等预处理, 之后将数据输入至PSAE-ATBiGRU网络威胁检测模型中进行训练。

(2) PSAE-ATBiGRU 网络威胁检测: 将数据测试集输入经过训练的威胁检测模型中, 根据模型输出结果记录各种攻击类型的发生次数以及误报消减矩阵, 用以计算网络安全态势值。

(3) 网络安全态势评估: 依据 PSAE-ATBiGRU 网络威胁检测模型的检测结果构建网络安全态势量化指标, 计算网络安全态势值并进行网络安全态势评估。

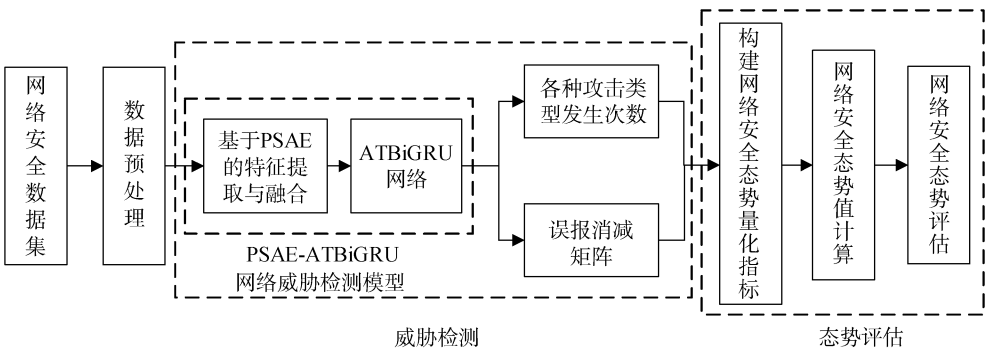

图 5　网络安全态势评估模型结构

Figure 5　Network security situation assessment framework

4.2　网络安全态势量化评估

网络安全态势评估结果通过影响网络安全的威胁严重度和威胁影响度确定。

(1) 威胁严重度

威胁严重度由各类攻击发生的次数、误报消减矩阵、各类攻击的威胁严重因子三项指标得出。其中, 各类攻击发生的次数、误报消减矩阵由 PSAE-BiGRU 模型测试的结果得到; 各类攻击的威胁严重因子在攻击威胁严重等级的基础上, 使用权系数生成法[35]计算得出。具体计算过程如下:

1) 获取各类攻击发生的次数

从测试数据集中随机选取若干组数据, 并将其输入到 PSAE-ATBiGRU 模型中, 对其进行攻击类型检测, 模型输出的各类攻击发生的次数为 C_i, 其中 i

代表攻击类型。

2) 获取误报消减矩阵

误报消减矩阵为 n 阶矩阵, 其中 n 代表模型测试结果的攻击类型个数。设数据集中 n 个攻击类型的下标集合为 $A=\{1, 2, \ldots, n\}$, a_{ij} 是模型测试结果为攻击类型 i 的样本个数中误报为攻击类型 j 的相对概率。将训练集输入训练完成的威胁检测模型中, 获得各种攻击类型发生的次数。根据模型测试结果与实际的攻击类型次数计算 a_{ij}, 得到模型的误报消减矩阵 $\boldsymbol{P}$

$$\boldsymbol{P}=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}。$$

然后, 计算各类攻击发生的修正次数 D_i

$$D_i=\begin{bmatrix} C_1 & C_2 & C_3 & \cdots & N_n \end{bmatrix}\cdot\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \end{bmatrix}^{\mathrm{T}} \tag{14}$$

3) 获取各类攻击的威胁严重因子

根据所采集的网络数据集中各类数据类型的主要攻击影响确定其威胁等级, 然后再使用权系数生成算法获取并计算各类攻击的威胁严重因子。本文采用的数据集为 NSL-KDD 数据集, 包括 4 种网络攻击类型和 1 种正常流量类型, 其基本情况如表 1 所示。

由于权系数生成算法[35]可在已知各类攻击的威胁等级的情况下, 计算各种攻击类型的威胁严重因子。所以, 在本文的评估方法中, 依据表 1 确定各种攻击类型的威胁等级, 再使用权系数生成算法计算威胁严重因子。设计具体处理过程如下:

按照攻击对网络的威胁程度可将 n 种攻击分为 $f(1\leqslant f\leqslant n)$个不同的威胁等级, 等级 k 的威胁严重因子 l_k

$$l_k=\begin{cases} \dfrac{1}{2}+\dfrac{\sqrt{-2\ln\dfrac{2k}{n}}}{6}, 1\leqslant k\leqslant\dfrac{n}{2} \\ \dfrac{1}{2}, k=\dfrac{n}{2} \\ \dfrac{1}{2}-\dfrac{\sqrt{-2\ln\dfrac{2k}{n}}}{6}, \dfrac{n}{2}<k\leqslant n \end{cases} \tag{15}$$

通过权系数生成算法得到各类攻击的威胁严重因子Q_i, 根据式(14)将各类攻击发生的次数C_i修正得到D_i。最后, 计算威胁严重度T_i

$$T_i=f(D_i,Q_i)=D_i\times 10^{Q_i} \tag{16}$$

表 1 5 种数据类型的主要攻击影响

Table 1 The main attack effects of the five data types

攻击类型	含义	主要攻击影响
Dos	拒绝服务攻击	耗尽资源、主机崩溃、无法使用资源、中断服务、无法通过或处理用户请求
U2R	获取最高权限控制主机	获取系统安全相关信息、获取用户密码与权限
R2L	远程攻击	获取高级用户权限、获取管理员权限、控制系统
Probe	监视和其他探测活动	扫描获取有关目标主机的信息、分析破坏目标主机安全性的因素
Normal	正常的流量	无影响

(2) 威胁影响度

机密性(confidentiality, C)度量攻击对信息资源的机密性的影响; 完整性(integrity, I)度量攻击对完整性造成的影响; 可用性(availability, A)度量攻击给受影响组件的性能带来的影响。通用漏洞评分系统(common vulnerability scoring system, CVSS)[36]中机密性、完整性、可用性的影响程度和分数如表 2 所示。

首先, 根据表 1 中各种攻击类型对机密性、完整性、可用性的影响程度进行等级划分并排序。

表 2 C、I、A 的影响分数

Table 2 Impact scores of C、I、A

指标	影响程度	分数
机密性 C	无(N)/低(L)/高(H)	0/0.22/0.56
完整性 I	无(N)/低(L)/高(H)	0/0.22/0.56
可用性 A	无(N)/低(L)/高(H)	0/0.22/0.56

然后, 结合表 2, 采用对数函数量化方法[37]计算得到各种攻击类型的威胁影响度 P_i

$$P_i=Round_2\left(\log_2\left(\frac{w_1 2^{Con_i}+w_2 2^{Int_i}+w_3 2^{Ava_i}}{3}\right)\right) \tag{17}$$

其中, Con_i、Int_i、Ava_i 分别指攻击类型 i 的 C、I、A 影响分数, w_1、w_2、w_3 分别对应 C、I、A 的权重。

(3) 网络安全态势值

首先, 计算得到网络安全态势值 R

$$R=\frac{1}{\left(N-C_n\right)}\sum_{i=1}^{n-1}T_i\times P_i \tag{18}$$

其中, N表示有N个样本, n表示有n种攻击类型, C_n表示Normal类型出现的次数。由于正常的网络流量对于网络环境无危害, 因此Normal类型的威胁严重度和威胁影响度为0, 只需计算n–1种攻击类型对网络安全态势的影响即可。

然后, 根据 R 值的区间, 参考《国家突发公共事件总体应急预案》[38]和 Snort 手册划分网络安全态势评估等级, 该安全态势评估等级包括: 安全、低危、中危、高危和超危 5 个等级, 对应的态势值区间和具体的说明如表 3 所示。

表 3 网络安全态势评估等级划分表

Table 3 Classification table of network security situation assessment

网络安全态势评估	态势值区间	说明
安全	0.00~0.30	威胁几乎不可能发生, 对网络环境无危害
低危	0.31~0.60	威胁出现的频率较小, 一般不太可能发生, 也没有被证实发生过, 对网络环境造成较小损害
中危	0.61~0.90	威胁出现的频率中等, 在某种情况下可能会发生或者被证实发生过, 对网络环境造成一般损害
高危	0.91~1.20	威胁出现的频率较高, 在大多数情况下很有可能会发生或者可以证实多次发生过, 对网络环境造成较大损害
超危	1.21~1.50	威胁出现的频率很高, 在大多数情况下几乎不可避免或者可以证实经常发生过, 对网络环境造成重大损害

5 实验与结果

为验证本文方法对网络安全态势评估的有效性和全面性, 通过实验验证PSAE特征提取器和注意力机制对基础模型 BiGRU 性能的提升效果。同时, 通过与典型方法的对比实验, 验证本文方法应用于网络安全态势评估的客观性与可行性。

上述实验均在 Ubuntu 系统上进行, 使用 TensorFlow 编程实现网络搭建, 并采用 TensorFlow-GPU[39]加速网络训练。实验配置为: Intel(R) Xeon(R) Silver 处理器、32GRAM、显卡为 RTX2060、内存 16G。

5.1 数据集描述与数据预处理

由于 NSL-KDD 数据集解决了 KDD99 数据集的故有问题[40], 其训练集 KDDTrain+不包含冗余记录、测试集 KDDTest+不包含重复记录、训练集和测试集记录数量设置合理, 故选取 NSL-KDD 数据集进行实验。NSL-KDD 数据集的基本信息如表 4 所示。

表 4 NSL-KDD 数据集信息

Table 4 NSL-KDD dataset information

数据集	Normal	Dos	Probe	R2L	U2R	Total
KDDTrain+	67343	45927	11656	995	52	125973
KDDTest+	9710	7458	2421	2754	200	22543

数据预处理过程包括特征数值化、特征约简、特征最大最小值归一化、平衡数据四项操作。

(1) 特征数值化

训练网络模型时需要将分类特征转化为连续值进行输入, NSL-KDD 数据集中包括三个分类特征, 因此, 通过独热编码(One-Hot)将其转化为分类向量来表示每个特征。例如, “protocol_type”的属性“tcp”、“udp”和“icmp”将分别转换为(1,0,1), (1,0,0)和(1,1,0)分类特征向量。用相同的方法将其余两个分类特征转化为对应的分类向量。完成所有转换后, 将数据集的特征维度从 41 个扩展为 122 个。

(2) 特征约简

NSL-KDD 数据集中有 15 个特征为零值, 由于它们的零值不会对模型训练结果产生影响且删除这些特征可以降低维度并提高训练效率, 因此, 删除这些冗余的特征, 将数据集的特征维度从 122 个缩减为 107 个。

(3) 特征最大最小归一化

NSL-KDD 数据集中部分特征的最大值和最小值之间的范围差异很大, 例如“duration”中最大值和最小值之间的差异最大为 58329, 最小为 0, “src-bytes”和“dst-bytes”等特征也存在较大差异。为消除特征之间单位和尺度差异对模型训练带来的影响,

应对特征进行归一化处理, 提升模型的训练效果。为此, 将特征映射至[0,1]区间

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{19}$$

其中, x 表示特征原始值, $x_{\min}$ 表示特征最小值, $x_{\max}$ 表示特征最大值。

(4) 平衡数据

从表 4 可见, NSL-KDD 数据集数据类型分布不平衡, 训练集 KDDTrain+中 Normal 类有 67343 条数据, 而 U2R 和 R2L 仅包含 52 和 995 条数据, 不同攻击类型数据量失衡会导致模型的弱检测问题。因此, 为了提高模型的检测效果, 本文采用 ADASYN 算法[41], 根据数据分布情况对不同类别的样本采样不同数量的新样本, 进而解决数据不平衡问题。

5.2 评价定义

为评估模型的性能, 实验选择正确分类为正常的样本数 *TN*(True Negatives)、错误分类为正常的攻击样本数 *FN*(False Negatives)、正确分类为攻击的样本数 *TP*(True Positives)、错误分类为攻击的正常样本数 *FP*(False Positives)用于定义以下指标:

准确率(Precision, *P*), 指学习模型正确预测为攻击的个数与学习模型预测为攻击的样本总数的百分比, 表示为

$$P = \frac{TP}{TP + FP} \times 100\% \tag{20}$$

召回率(Recall, *R*), 指学习模型正确预测为攻击的个数与真实类别为攻击的样本总数的百分比, 表示为

$$R = \frac{TP}{TP + FN} \times 100\% \tag{21}$$

*F*1 值(F1-score, *F*1), 综合考虑了 *P* 和 *R*, 是衡量模型检测性能的重要指标, 表示为

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \tag{22}$$

5.3 模型训练与模型检测

实验选取训练集 KDDTrain+中的 125973 条数据作为训练集进行学习, 预训练学习率为 le-3, 当准确率 20 轮内不再提升时, 将学习率减少为原来的 0.5 倍, 每一批次输入 1024 条数据, 网络迭代训练 200 次。训练完成后, 选取测试集 KDDTest+中的 22543 条数据作为测试集进行威胁检测。

为了分析本文所提模型 PSAE-ATBiGRU 的威胁检测准确率, 与原始模型 BiGRU、仅用 PSAE 对原始模型进行改进的模型 PSAE-BiGRU 和仅用注意力机制改进的模型 ATBiGRU 进行对比, 图 6 展示了训练过程中测试集在 4 种模型上的准确率变化情况。

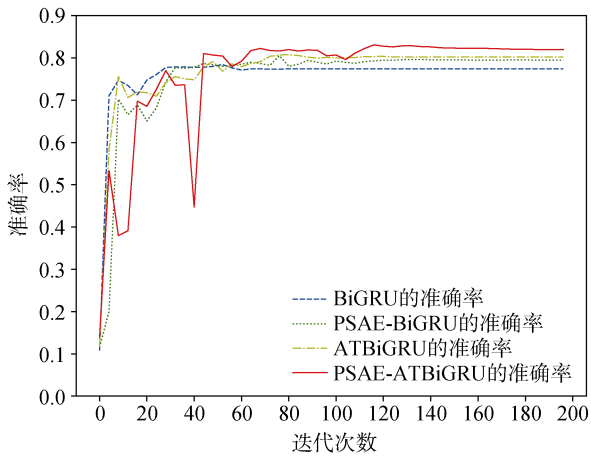

图 6 4 种模型的威胁检测准确率

Figure 6 Threat detection accuracy of four models

首先, 从图 6 我们可以看到, 在训练过程中, 迭代次数为 40 次附近时模型的准确率波动较大, 但后期准确率趋于稳定。这是由于我们在训练过程中, 采用了动态的学习率调整策略, 训练早期学习率较大, 模型还未很好的拟合数据的分布, 导致模型在最优解附近震荡。训练后期, 模型已经可以较好的拟合数据分布, 此时学习率动态调整到较小的值, 准确率趋于稳定。

其次, 由图 6 可见, 与 BiGRU 模型相比, PSAE-BiGRU 和 ATBiGRU 两种模型的准确率分别提高了 2.85%和 3.64%, 本文模型的准确率为 82.13%, 比 BiGRU 模型提高了 5.28%。原因在于本文模型采用 PSAE 提高原始数据的表征能力, 通过注意力机制进行加权特征学习, 突出了上述两种方法的优点。

分别从准确率、召回率和 *F*1 值方面比较分析上述 4 种模型, 实验结果见图 7。其中, 纵坐标表示模型评价得分, 数字越大表明模型性能越好。对比结果表明, 本文模型的准确率、召回率、*F*1 值均优于其他 3 个模型。与 BiGRU、PSAE-BiGRU 和 ATBiGRU 模型相比, 本文模型的准确率分别提高了 5.28%、2.43%、1.64%; 召回率分别提高了 5.65%、2.58%、1.42%; *F*1 值分别提高了 5.46%、2.5%和 1.53%。

5.4 网络安全态势评估结果与分析

为评估网络的整体态势, 须对影响网络安全的威胁严重度和威胁影响度两个影响因素进行量化评估。首先, 通过网络威胁测试获取各类攻击发生的次数和误报消减矩阵, 再结合各类攻击的威胁严重因子确定威胁严重度。然后, 结合 4.2 节各类攻击的威胁影响度计算网络安全态势值。最后, 依据态势值区间对照表 3 确定网络的整体安全态势评估结果。

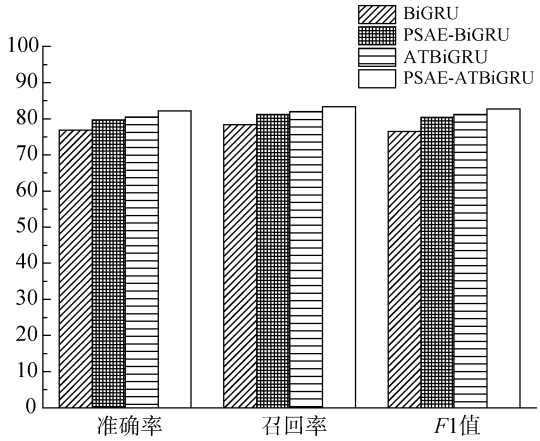

图 7　4 种模型的准确率、召回率、*F*1 值

Figure 7　Accuracy, recall, and *F*1 of four models

随机从测试集中选取 100 组相同数据数量的测试样本集合。将其作为输入数据对 BiGRU、PSAE-BiGRU、ATBiGRU 和 PSAE-ATBiGRU 4 种模型进行 100 组测试实验, 采用本文态势值量化方法得到基于上述 4 种模型的网络安全态势值, 结合网络的实际态势值计算每种模型的网络安全态势值测试误差值。通过将式(16)中的 D_i 替换为测试样本中各种攻击类型的实际次数, 由式(17)、(18)计算得到实际态势值。图 8 展示了其中 20 组的归一化态势值测试误差值 λ。

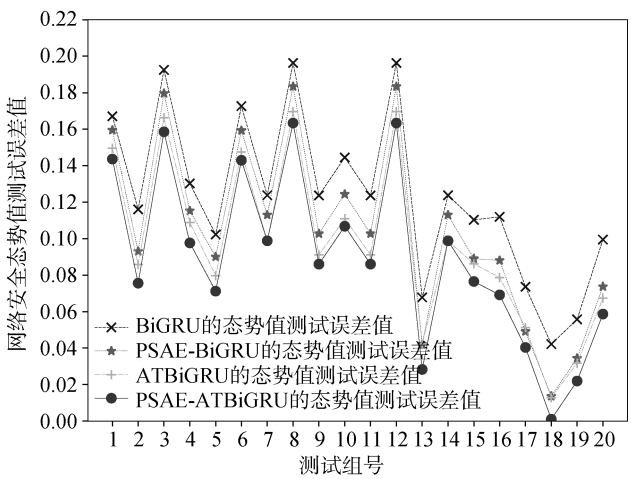

图 8　4 种模型的网络安全态势测试误差

Figure 8　The network security situation test errors of four models

由图 8 可见, BiGRU 模型的误差值最大, 而在此模型上改进的 PSAE-BiGRU 和 ATBiGRU 模型的误差值均小于 BiGRU 模型, 验证了本文方法的有效性。与 3 种模型相比, 基于本文模型 PSAE-BiGRU 得到网络安全态势值与真实值的测试误差值 λ 最小, 这说明本文方法对网络安全威胁的检测能力更突出, 计算出的网络安全态势值更符合实际的网络安全态势情况。

为进一步验证评估结果的客观性与真实性, 从 NSL-KDD 测试集中随机选取相同数量的测试样本, 采用 SVM[25]、LSTM[21]、BiGRU[20]、AEDNN[22]、PSAE-ATBiGRU 模型进行威胁检测实验。根据威胁检测结果获取每个模型在每组测试实验中各类攻击发生的次数。最后, 结合每个模型的误报消减矩阵、各类攻击的威胁严重因子、各类攻击的 C、I、A 影响分数, 采用 4.2 节态势值计算方法得到基于上述 5 种模型的网络安全态势值。图 9 展示了其中 20 组实验的网络态势值对比结果。

由图 9 可见, PSAE-ATBiGRU 模型得到的网络安全态势值和真实的态势值始终位于同一态势评估区间, 而 SVM、LSTM、BiGRU 和 AEDNN 模型得到的态势值存在与真实态势值不在同一区间的情况。如: 在第 2、15 组中, SVM、LSTM、BiGRU 和 AEDNN 模型的网络安全态势评估结果为中危, 而真实的态势情况为低危; 在第 3 组中, SVM、LSTM、BiGRU 和 AEDNN 模型的网络安全态势评估结果为中危, 而真实的态势情况为高危。这表明, PSAE-ATBiGRU 模型的态势评估结果更贴合实际的网络态势情况。

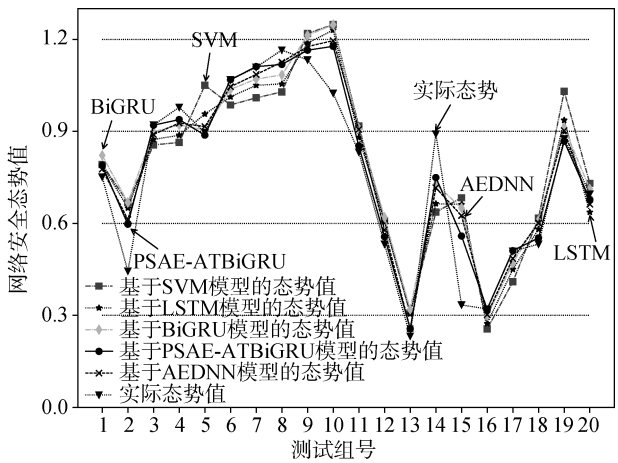

图 9　5 种模型的网络安全态势值对比

Figure 9　Comparison of network security situation values of five models

此外, 图 9 的部分测试结果中, SVM、LSTM、BiGRU、AEDNN 和 PSAE-ATBiGRU 模型的态势值均与真实的态势值在同一态势评估区间, 但是, PSAE-ATBiGRU 模型得到的网络安全态势值始终与真实的态势值更接近。如: 在第 1、6 组中, 5 个模型的态势值与真实的态势值均在同一态势评估区间, 但是 PSAE-ATBiGRU 模型的态势值与真实态势值之间的误差更小。这表明, PSAE-ATBiGRU 模型对网络

威胁的表征能力更强。

从测试数据集中随机选取 10 组相同数量的测试样本, 模拟某一时间段内网络受到的威胁攻击情况并进行测试实验。在 10 个相同时间段内, 分别采用 SVM、LSTM、BiGRU、AEDNN 和 PSAE-ATBiGRU 模型计算网络安全态势值与实际安全态势值对比误差, 然后计算每段时间内 5 种模型的均方根误差值。由表 5 可见, AEDNN 模型的均方根误差值小于 SVM、LSTM 和 BiGRU, 因为该模型应用 UOSW 算法[22]提高了 U2R 和 R2L 两种少训练样本类别的准确率。此外, PSAE-ATBiGRU 模型的均方根误差值最小, 其学习结果优于其他 4 种模型, 由该模型得到的安全态势值与真实安全态势值最接近, 其检测效果更符合实际。

表 5 5 种模型的均方根误差值

Table 5 Root mean square errors of five models

模型	均方根误差值
SVM	0.2915
LSTM	0.2328
BiGRU	0.1961
AEDNN	0.1768
PSAE-ATBiGRU	0.1011

表 6 具体展示了由本文方法得到的 10 个时间段内的安全态势评估结果与实际态势情况。由表 6 可见, 本文方法计算的态势值与实际态势值之间存在些许差异, 但评估结果落在了相同的区域, 根据表 3 定义的网络安全态势等级, 本文方法的态势评估结果与实际情况相符。

表 6 态势值和网络安全态势评估情况

Table 6 Situation value and network security situation assessment

时间段/min	PSAE-ATBiGRU		实际情况	
	态势值	态势等级	态势值	态势等级
0~10	0.5973	低危	0.4443	低危
11~20	0.7889	中危	0.7507	中危
21~30	0.9201	高危	0.9203	高危
31~40	0.9385	高危	0.9784	高危
41~50	0.8141	中危	0.7708	中危
51~60	1.0685	高危	1.0680	高危
61~70	1.1111	高危	1.1089	高危
71~80	0.5586	低危	0.3341	低危
81~90	1.1182	高危	1.1652	高危
91~100	0.7495	中危	0.8915	中危

6 结论

本文提出了一种基于深度加权特征学习的网络安全态势评估方法。该方法使用并行特征提取方法有效增强提取特征对原始数据的表征能力, 应用注意力机制对 BiGRU 网络进行改进从而确定不同特征的最佳权重。通过 PSAE-ATBiGRU 对网络威胁进行检测并根据检测结果以及误报消减矩阵评估网络安全态势。通过与 BiGRU、LSTM、SVM、AEDNN 等方法的评估对比实验, 表明本文方法获得的网络安全态势评估结果的有效性和可靠性更具优势。

在未来的研究中, 拟考虑将本文模型应用于更多种类的网络安全数据集的威胁检测。除此之外, 研究更加有效的优化算法以提高模型建模速度, 进一步减少模型的训练和测试时间。

参考文献

[1] China Internet Network Information Center. The 47th China Statistical Report on Internet Development [EB/OL]. http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm, 2021.
(中国互联网络信息中心. 第 47 次中国互联网络发展状况统计报告[EB/OL]. http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm, 2021.)

[2] National Internet Emergency Center. Summary of Internet Network Security Situation in China in 2020[EB/OL]. https://www.cert.org.cn/publish/main/upload/File/2020%20CNCERT%20Cybersecurity%20Analysis.pdf, 2021.
(国家互联网应急中心. 2020 年我国互联网网络安全态势综述[EB/OL]. https://www.cert.org.cn/publish/main/upload/File/2020%20CNCERT%20Cybersecurity%20Analysis.pdf, 2021.)

[3] Bass T. Intrusion Detection Systems and Multisensor Data Fusion[J]. *Communications of the ACM*, 2000, 43(4): 99-105.

[4] Zhao D M, Liu J X. Study on Network Security Situation Awareness Based on Particle Swarm Optimization Algorithm[J]. *Computers & Industrial Engineering*, 2018, 125: 764-775.

[5] Gong J, Zang X D, Su Q, et al. Survey of Network Security Situation Awareness[J]. *Journal of Software*, 2017, 28(4): 1010-1026.
(龚俭, 臧小东, 苏琪, 等. 网络安全态势感知综述[J]. *软件学报*, 2017, 28(4): 1010-1026.)

[6] Lu S, Zhuang Y. A Network Security Situational Awareness Framework Based on Situation Fusion[C]. *Security*, *Privacy*, *and Anonymity in Computation*, *Communication*, *and Storage*, 2021: 345-355.

[7] Agrawal A, Seh A H, Baz A, et al. Software Security Estimation Using the Hybrid Fuzzy ANP-TOPSIS Approach: Design Tactics Perspective[J]. *Symmetry*, 2020, 12(4): 598.

[8] Wen L B. Security Evaluation of Computer Network Based on Hierarchy[J]. *International Journal of Network Security*, 2019, 21(5): 735-740.

[9] Li J, Yan L N, Wang J Y, et al. Research on Network Security Risk Assessment Method Based on Improved AHP[J]. *Journal of Phys-*

ics: *Conference Series*, 2021, 1828(1): 012115.

[10] Han M N, Liu Y, Chen Y. Network Security Situational Awareness Model Based on Set Pair Analysis[J]. *Application Research of Computers*, 2012, 29(10): 3824-3827.
(韩敏娜, 刘渊, 陈烨. 基于集对分析的网络安全态势评估[J]. *计算机应用研究*, 2012, 29(10): 3824-3827.)

[11] Huang J Z. Research on Model of Network Security Evaluation Based on Fuzzy Mathematics[J]. *Cyberspace Security*, 2020, 11(4): 1-4.
(黄加增. 基于模糊数学的网络安全评价模型研究[J]. *网络空间安全*, 2020, 11(4): 1-4.)

[12] Alali M, Almogren A, Hassan M M, et al. Improving Risk Assessment Model of Cyber Security Using Fuzzy Logic Inference System[J]. *Computers & Security*, 2018, 74: 323-339.

[13] Yang H Y, Zhang X G. Self-Corrected Coefficient Smoothing Method Based Network Security Situation Prediction[J]. *Journal on Communications*, 2020, 41(5): 196-204.
(杨宏宇, 张旭高. 基于自修正系数修匀法的网络安全态势预测[J]. *通信学报*, 2020, 41(5): 196-204.)

[14] Pu Z Y. Network Security Situation Analysis Based on a Dynamic Bayesian Network and Phase Space Reconstruction[J]. *The Journal of Supercomputing*, 2020, 76(2): 1342-1357.

[15] Li X N, Li M G, Wang H. Research on Network Security Risk Assessment Method Based on Bayesian Reasoning[C]. *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication*, 2019: 1-7.

[16] Yang H Y, Feng Y H. A Pythagorean Fuzzy Petri Net Based Security Assessment Model for Civil Aviation Airport Security Inspection Information System[J]. *International Journal of Intelligent Systems*, 2021, 36(5): 2122-2143.

[17] Zhao Z W, Peng Y, Huang J H, et al. An Evaluation Method of Network Security Situation Using Data Fusion Theory[J]. *International Journal of Performability Engineering*, 2020, 16(7): 1046.

[18] Yang H Y, Wang F Y. Network Threat Situation Assessment Based on Unsupervised Multi-Source Data Feature Analysis[J]. *Journal on Communications*, 2020, 41(2): 143-154.
(杨宏宇, 王峰岩. 基于无监督多源数据特征解析的网络威胁态势评估[J]. *通信学报*, 2020, 41(2): 143-154.)

[19] Song Y. Evaluation of Network Security Situation Based on Grey Relational Analysis and Support Vector Machine[J]. *Laser Journal*, 2015, 36(4): 147-150.
(宋严. 灰色关联分析与支持向量机相融合的网络安全态势评价[J]. *激光杂志*, 2015, 36(4): 147-150.)

[20] Lin Y, Wang J, Tu Y, et al. Time-Related Network Intrusion Detection Model: A Deep Learning Method[C]. *2019 IEEE Global Communications Conference*, 2019: 1-6.

[21] Li S X, Zhao D M. A LSTM-Based Method for Comprehension and Evaluation of Network Security Situation[C]. *2019 18th IEEE International Conference on Trust*, *Security and Privacy In Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering*, 2019: 723-728.

[22] Yang H Y, Zeng R Y, Xu G Q, et al. A Network Security Situation Assessment Method Based on Adversarial Deep Learning[J]. *Applied Soft Computing*, 2021, 102: 107096.

[23] Sreenivasa Chakravarthi S, Jagadeesh R. Non-Linear Dimensionality Reduction-Based Intrusion Detection Using Deep Autoencoder[J]. *International Journal of Advanced Computer Science and Applications*, 2019, 10(8): 168-174.

[24] Vartouni A M, Kashi S S, Teshnehlab M. An Anomaly Detection Method to Detect Web Attacks Using Stacked Auto-Encoder[C]. *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems*, 2018: 131-134.

[25] Hu J J, Ma D Y, Liu C, et al. Network Security Situation Prediction Based on MR-SVM[J]. *IEEE Access*, 2019, 7: 130937-130945.

[26] Shone N, Ngoc T N, Phai V D, et al. A Deep Learning Approach to Network Intrusion Detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41-50.

[27] Liu T L, Qi Y, Shi L, et al. Locate-then-Detect: Real-Time Web Attack Detection via Attention-Based Deep Neural Networks[C]. *The Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019: 4725-4731.

[28] Kundu A, Sahu A, Serpedin E, et al. A3D: Attention-Based Auto-Encoder Anomaly Detector for False Data Injection Attacks[J]. *Electric Power Systems Research*, 2020, 189: 106795.

[29] Yang S C, Tan M S, Xia S Y, et al. A Method of Intrusion Detection Based on Attention-LSTM Neural Network[C]. *The 2020 5th International Conference on Machine Learning Technologies*, 2020: 46-50.

[30] Tavallaee M, Bagheri E, Lu W, et al. A Detailed Analysis of the KDD CUP 99 Data Set[C]. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009: 1-6.

[31] Javaid A, Niyaz Q, Sun W Q, et al. A Deep Learning Approach for Network Intrusion Detection System[C]. *The 9th EAI International Conference on Bio-inspired Information and Communications Technologies* (*formerly BIONETICS*), 2016: 21-26.

[32] Ma C, Yang C S, Yang F, et al. Trajectory Factory: Tracklet Cleaving and re-Connection by Deep Siamese Bi-GRU for Multiple Object Tracking[C]. *2018 IEEE International Conference on Multimedia and Expo*, 2018: 1-6.

[33] Luong T, Pham H, Manning C D. Effective Approaches to Attention-Based Neural Machine Translation [EB/OL]. 2015: ArXiv Preprint ArXiv:1508.04025.

[34] Yang Z C, Yang D Y, Dyer C, et al. Hierarchical Attention Networks for Document Classification[C]. *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, 2016: 1480-1489.

[35] Liu X W, Wang H Q, Lü H W, et al. Fusion-Based Cognitive Awareness-Control Model for Network Security Situation[J]. *Journal of Software*, 2016, 27(8): 2099-2114.
(刘效武, 王慧强, 吕宏武, 等. 网络安全态势认知融合感控模型[J]. *软件学报*, 2016, 27(8): 2099-2114.)

[36] Common Vulnerability Scoring System v3.0: Specification Document, last accessed 2020/06/22.https://www.first.org/cvss/ specification-document.

[37] Xi R R, Yun X C, Zhang Y Z. Quantitative Threat Situational Assessment Based on Contextual Information[J]. *Journal of Software*, 2015, 26(7): 1638-1649.

(席荣荣, 云晓春, 张永铮. 基于环境属性的网络威胁态势量化评估方法[J]. *软件学报*, 2015, 26(7): 1638-1649.)

[38] The State Council of the People's Republic of China. Overall Emergency Plans for National Sudden Public Incidents[M]. BEIJING: China Legal Press, 2006.
(国务院. 国家突发公共事件总体应急预案[M]. 北京: 中国法制出版社, 2006.)

[39] TensorFlow-GPU [[EB/OL]. https://tensorflow.google.cn, 2021.

[40] Ferrag M A, Maglaras L, Moschoyiannis S, et al. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study[J]. *Journal of Information Security and Applications*, 2020, 50: 102419.

[41] Yang L Q, Zhang J W, Wang X Z, et al. An Improved ELM-Based and Data Preprocessing Integrated Approach for Phishing Detection Considering Comprehensive Features[J]. *Expert Systems With Applications*, 2021, 165: 113863.

杨宏宇 于 2003 年在天津大学计算机应用技术专业获得博士学位。现任中国民航大学安全科学与技术学院、计算机科学与技术学院教授。研究领域为网络与系统安全。Email: hyyang@cauc.edu.cn

张良 于 2017 年在天津大学信息与通信工程专业获得博士学位。现为亚利桑那大学博士后研究员。研究领域为强化学习和基于深度学习的信号处理。Email: liangzh@arizona.edu

张梓锌 于 2019 年在中国民航大学信息安全专业获得学士学位。现在中国民航大学计算机技术专业攻读硕士学位。研究领域为网络与系统安全。Email: zixin_zhang2021@163.com