DOI:10.19651/j.cnki.emt.2211448

一种基于 IDOA-RBF 神经网络的正常流量过滤方法*

钱 来 王 伟
(西安工程大学计算机科学学院 西安 710048)

摘 要: 针对全流量检测方式容易使安全检测设备出现性能瓶颈的问题,给出一种使用改进的野狗优化算法来优化径向基函数神经网络的正常流量过滤方法。首先,采用 Singer 混沌映射和搜索平衡策略对野狗优化算法进行改进;其次,用改进后的野狗优化算法优化 RBF 神经网络的输出权值,使用 CSE-CIC-IDS2018 数据集训练网络,构建正常流量过滤模型;最后,在网络流量进入安全检测设备前尽可能多地过滤掉其中正常流量,减轻安全检测设备的工作负担。实验结果表明:与现有的模型相比,IDOA-RBF 神经网络的正常流量过滤模型在建模时间上有较大的改善,同时保持较高的识别精度,并且能在需要检测的流量中过滤掉 72.9%的正常流量。

关键词: 流量识别;流量过滤;野狗优化算法;径向基函数(RBF)神经网络;CSE-CIC-IDS2018 数据集
中图分类号: TP 301.6 **文献标识码:** A **国家标准学科分类代码:** 510.4030

A normal traffic filtering method based on IDOA-RBF neural network

Qian Lai Wang Wei
(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Aiming at the problem that the full flow detection mode is easy to cause the performance bottleneck of the security detection equipment, an improved dingo optimization algorithm is given to optimize the radial basis function RBF neural network for normal traffic filtering. First, the wild dog optimization algorithm was improved using Singer chaotic mapping and search balance strategy; second, the output weights of the RBF neural network were optimized with the improved wild dog optimization algorithm, and the network was trained using the CSE-CIC-IDS2018 dataset to construct a normal traffic filtering model. Finally, before the network traffic entered the security detection device, filter out as many normal traffic as possible to reduce the workload of the security detection device. The experimental results show that compared with the existing models, the normal traffic filtering model of IDOA-RBF neural network has a great improvement in modeling time, while maintaining a high recognition accuracy, and can filter out 72.9% of the normal traffic in the traffic to be detected.

Keywords: traffic identification; traffic filtering; dingo optimization algorithm; radial basis function neural network; CSE-CIC-IDS2018 data set

0 引 言

当今世界,由于网络攻击类型的复杂性[1],在大量的数据流量中精准识别出异常的流量变得越来越困难[2]。通常,基于神经网络的异常数据流量检测是利用网络数据流量的特征对其进行识别分类,具有较高的异常数据识别精度和检测效率,具有广阔的应用前景[3-4]。常采用网络模型有反向传播(back propagation,BP)神经网络模型[5]、深度学习[6-7]、极限学习机[8-9]、自编码器[10]等。上述神经网络模型存在学习规则复杂,收敛速度较慢,容易陷入局部最优等问题。改进后的网络模型在检测精度方面有所提高,但是整体构建网络模型时网络收敛慢,时间开销大。所以,需要寻找一种学习规则简单、收敛性能好的神经网络。

径向基函数(radial basis function,RBF)神经网络是一种有局部函数逼近能力的神经网络,具有克服局部最小值,收敛性能优异,学习规则容易实现等特点。解男男提出利用粒子群算法的寻优特点去计算 RBF 神经网络的最优权值,创建入侵检测网络模型,提高了网络模型的收敛速

收稿日期:2022-09-19
* 基金项目:2021 年中国高校产学研创新基金(2021ALA02002)、2021 年"纺织之光"中国纺织工业联合会高等教育教学改革研究项目(2021BKJGLX004)、西安工程大学 2020 年高等教育研究项目(20GJ05)、陕西省自然科学基础研究计划(2019JM-291)项目资助

度[11]。吴贻淮通过遗传算法优化传统 RBF 神经网络的初始权重,实验结果表明此改进方案提高了 RBF 神经网络对未知网络数据的识别能力[12]。但是这些入侵检测使用的检测方式是对全部的网络流量进行安全检测,实际网络场景下,异常网络攻击流量一般是全部网络流量的千分之一,安全检测设备的大量解包动作会造成资源的浪费。为了提高安全检测设备的准确度,需要识别并过滤流量中大部分正常流量。

本文利用对少量的数据流量进行检测时速度快,准确率高的特点[13],提出一种正常流量过滤方法,通过对野狗优化算法(dingo optimization algorithm,DOA)进行改进,提升算法的求解精度,然后使用改进的野狗优化算法(improved dingo optimization algorithm,IDOA)优化 RBF 神经网络,即 IDOA-RBF。让模型具有自适应训练参数能力的同时提高模型的分类识别能力。在进行流量深度解析前识别并过滤掉其中大部分正常流量,减少安全检测设备深度解析数据包的次数,单位时间内对更多可能存在威胁的流量进行分析,提高安全检测设备的性能。

1　正常流量过滤模型

1.1　RBF 神经网络

RBF 神经网络由 3 层网络结构组成,相邻的 2 个层均是单向链接的。训练数据与测试数据通过输入层单元使用非线性函数映射到隐藏层,隐含层再通过神经元使用 RBF 传送到输出层,然后由输出层给出结果[14]。具有一层隐藏层的 RBF 神经网络,如图 1 所示。

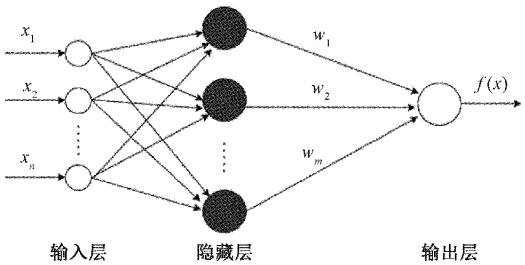

图 1　RBF 神经网络

输入层与输出层的函数映射关系如式(1)所示。

$$f(x)=\sum_{i=1}^{m}w_i exp(-\frac{1}{2\sigma^2}\parallel x-c_i\parallel^2) \tag{1}$$

式中:c_i 为径向基函数的中心;w_i 为权值;σ 为扩展宽带,其值越小,则表明基函数选择性越好。

1.2　野狗优化算法

自然从一开始就是强大的老师,曾经在地球上生存的物种都有其独特的生存机制。为解决复杂繁琐的问题人类对不同物种进行研究,获得新的解决思路[15]。DOA 就是一种新颖的仿生算法,用于模拟野狗狩猎策略的全局优化。该策略通过群攻、迫害和清道夫等行为对目标发动攻击。由于澳大利亚野狗濒临灭绝,所以在算法中将野狗的生存概率考虑在内[16]。

DOA 数学模型如下:

策略 1:群攻。野狗在捕猎时会成群结队,快速找到猎物的位置并将其包围。如式(2)所示。

$$\boldsymbol{x}_i(t+1)=\beta_1\sum_{k=1}^{n}\frac{[\boldsymbol{\varphi}_k(t)-\boldsymbol{x}_i(t)]}{n}-\boldsymbol{x}_\theta(t) \tag{2}$$

式中:$\boldsymbol{x}(t+1)$ 为搜索代理的新位置;n 为一个在 $[2,S/2]$ 之间的随机整数,其中 S 为野狗的总数;$\boldsymbol{\varphi}_k(t)$ 为搜索代理的子集,其中 $\boldsymbol{\varphi}\subset\boldsymbol{X}$;$\boldsymbol{X}$ 为随机产生的野狗种群;$\boldsymbol{x}_i(t)$ 为当前的搜索代理;$\boldsymbol{x}_\theta(t)$ 为从上一次迭代中找到的最佳搜索代理;β_1 为区间$[-2,2]$内均匀生成的随机数,它是改变野狗轨迹大小和意义的比例因子。

策略 2:迫害。野狗通常捕猎小猎物,并一直追赶,直到猎物被捕获。如式(3)所示。

$$\boldsymbol{x}_i(t+1)=\boldsymbol{x}_\theta(t)+\beta_1 e^{\beta_2}(\boldsymbol{x}_{r_1}(t)-\boldsymbol{x}_i(t)) \tag{3}$$

式中:$\boldsymbol{x}_i(t+1)$ 为野狗的运动;$\boldsymbol{x}_i(t)$ 为当前的搜索代理;$\boldsymbol{x}_\theta(t)$ 为从上一次迭代中找到的最佳搜索代理;β_1 与式(2)中的值相同;β_2 为区间$[-1,1]$内均匀生成的随机数;r_1 为区间$[1,\boldsymbol{x}_i(t)_{\max}]$内生成的随机数,$\boldsymbol{x}_i(t)_{\max}$ 为最大搜索代理;$\boldsymbol{x}_{r_1}(t)$ 为第 r_1 个搜索代理,其中 $i\neq r_i$。

策略 3:清道夫。清道夫行为是当野狗在栖息地内行走发现腐肉时的行为。如式(4)所示。

$$\boldsymbol{x}_i(t+1)=\frac{1}{2}[e^{\beta_2}\boldsymbol{x}_{r_1}(t)-(-1)^\sigma\boldsymbol{x}_i(t)] \tag{4}$$

式中:$\boldsymbol{x}_i(t+1)$ 为野狗的运动;$\boldsymbol{x}_i(t)$ 为当前搜索代理;β_2 与式(3)中的值相同;r_1 为区间$[1,\boldsymbol{x}_i(t)_{\max}]$内生成的随机数,$\boldsymbol{x}_i(t)_{\max}$ 为最大搜索代理;$\boldsymbol{x}_{r_1}(t)$ 为第 r_1 个搜索代理,其中 $i\neq r_i$;σ 是随机生成的二进制数。

策略 4:野狗的存活率。如式(5)所示。

$$s_i=\frac{f_{\max}-f_i}{f_{\max}-f_{\min}} \tag{5}$$

式中:f_{max} 和 f_{min} 为当前一代中最佳和最差的适应度值;f_i 为第 i 个搜索代理的适应度值;s_i 为归一化的野狗存活率。

低存活率情况时的搜索策略。如式(6)所示。

$$\boldsymbol{x}_i(t)=\boldsymbol{x}_\theta(t)+\frac{1}{2}[\boldsymbol{x}_{r_1}(t)-(-1)^\sigma\boldsymbol{x}_{r_2}(t)] \tag{6}$$

式中:$\boldsymbol{x}_i(t)$ 为具有低存活率的搜索代理;r_1 和 r_2 为区间$[1,\boldsymbol{x}_i(t)_{\max}]$内生成的随机数,$\boldsymbol{x}_i(t)_{\max}$ 为最大搜索代理,其中 $r_1\neq r_2$;$\boldsymbol{x}_{r_1}(t)$ 和 $\boldsymbol{x}_{r_2}(t)$ 为选择第 r_1 和 r_2 个搜索代理;$\boldsymbol{x}_\theta(t)$ 为从上一次迭代中找到的最佳搜索代理;σ 是随机生成的二进制数。

1.3　改进的野狗优化算法

1) Singer 混沌映射初始化

在原始的 DOA 中,野狗的位置是通过随机初始化产生的,这就出现野狗的位置分布过于零散,降低了算法的寻优能力。而 Singer 映射是混沌映射的一种典型方式,其具有遍历性和随机性的特点,初始化的种群更加均匀[17]。因

此使用 Singer 映射对野狗种群初始化。

$$a = 7.86, b = 23.31, c = 28.75, d = 13.302\ 875$$

$$\boldsymbol{x}_i(t) = u(a\boldsymbol{x}_\theta(t) - b\boldsymbol{x}_\theta(t)^2 + c\boldsymbol{x}_\theta(t)^3 - d\boldsymbol{x}_\theta(t)^4) \tag{7}$$

式中：$\boldsymbol{x}_i(t)$ 为当前的搜索代理；$\boldsymbol{x}_\theta(t)$ 为从上一次迭代中找到的最佳搜索代理；$u \in (0.9, 1.08)$，本文中 $u = 1.07$。

2）搜索平衡策略

由于仿生算法需要在局部搜索和全局搜索之间取得良好的平衡[18]。本文在生存策略中加入局部搜索程序，寻找其他最佳方案替换迄今为止获得的方案。因此，使用式(6)与全局搜索相关联，式(8)与局部搜索相关联。

$$\boldsymbol{x}_i(t) = \boldsymbol{x}_\theta(t) + \boldsymbol{\alpha}\left(\frac{1}{2} - \delta\right) \tag{8}$$

式中：$\boldsymbol{x}_i(t)$ 为具有低存活率的搜索代理；$\boldsymbol{x}_\theta(t)$ 为从上一次迭代中找到的最佳搜索代理；$\boldsymbol{\alpha}$ 为区间(−2,2)内均匀分布的伪随机数；δ 为区间(0,1)内正态分布的伪随机数。

如图 2 所示，当索引号对应的 s 小于 0.3，并且索引号是偶数，将应用全局搜索；对于索引号是奇数，将应用局部搜索。

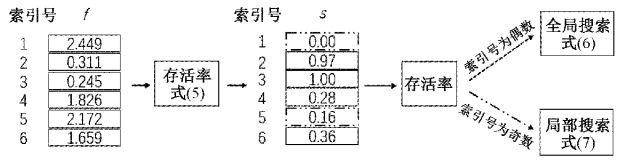

图 2　局部搜索与全局搜索使用方案

1.4　适应度函数

启发式优化算法在求解问题时，会把问题映射成目标函数，再将目标函数充当启发式优化算法的适应度函数进行求解[19]。本文采用 IDOA 对目标函数求最优解，使用 IDOA-RBF 神经网络模型识别结果的均方误差(mean square error，MSE)作为适应度函数如式(9)所示。

$$EMS = \frac{1}{m}\sum_{i=1}^{m}(y_i^k - y_i^k)^2 \tag{9}$$

式中：y_i^k 为样本 k 的第 i 个识别数据值；$\bar{y}_i^k$ 为样本 k 的第 i 个真实数据值；m 为样本数据量。

1.5　IDOA-RBF 算法

在网络数据流量识别模型训练中，RBF 神经网络相对于 BP 神经网络有着结构简单，训练速度快，输出结果具有全局最优的优点。但是网络模型神经元的增多导致最优输出权值求解过程更为复杂，构建 RBF 神经网络模型时，神经网络的输出权值很难确定[20]。因此本文提出一种 IDOA 对 RBF 神经网络权值进行优化，构建基于 IDOA-RBF 神经网络的正常流量过滤模型。IDOA-RBF 算法流程如图 3 所示。

输入数据集训练 RBF 网络模型的基函数中心 c_i 和扩展宽带 σ，把需要优化的权值映射为 IDOA 中野狗个体；对野狗的位置进行初始化，产生搜索代理；将野狗个体重新映射为权值；计算适应度，更新搜索代理，直至满足最大迭代次数或精度；输出最优代理，映射为权值，建立新的 RBF 网络模型；RBF 网络对流量进行识别，并过滤正常流量。

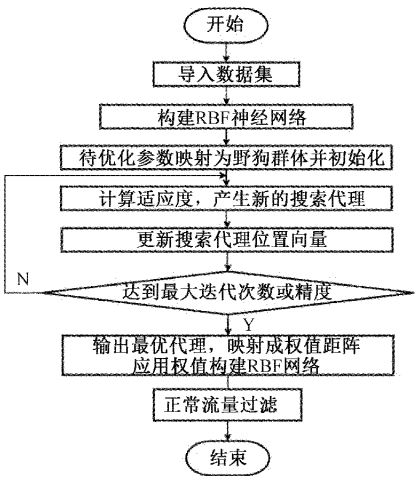

图 3　IDOA-RBF 算法流程

2　实验数据集

2.1　CSE-CIC-IDS2018 数据集

理想的安全数据集应包含最新的网络攻击和正常的上网流量数据，需要正常标记、公开可用和数据量足够[21]。但是在网络安全实验环境中，数据集只能做到相对理想。目前主流的网络安全实验数据集有 KDD-Cup99 与 NSL-KDD 等，但这些数据集包含的攻击流量有些老旧，没有采集到目前新的网络攻击流量[22]。CSE-CIC-IDS2018 是通信安全机构和加拿大网络安全研究所共同采集创建的数据体量庞大的入侵检测数据集。研究表明该数据集包括当时最新的网络攻击数据，而且其攻击数据集的标准也符合现实网络环境，在理论上没有较大的问题[23]。所以，本文使用 CSE-CIC-IDS2018 数据集对模型进行训练和验证。

2.2　数据集预处理

通过对 CSE-CIC-IDS2018 数据集进行分析，发现其中存在大量的异常数据值。所以需要对数据集中的异常值进行修改和删除，提升其质量。CSE-CIC-IDS2018 数据集一共有 10 个不同时间段的流量数据子文件，由于每个文件数据量比较大，本文首先对每个子文件进行数据预处理，最后将处理好的子文件合并成所需的数据集。

1）异常值处理

分析数据集后发现在 Flow Byts/s 特征下有缺失值存在，在数据集中占的比例较少，因此对此特征进行删除；在特征 Flow Byts/s 和 Flow Pkts/s 下有无法参与计算的无穷值，且无穷值分布在流量条目最多的正常流量上，因此将这类特征进行删除；数据集中还存在重复数据，数据重复会增加网络模型计算开销，因此只保留重复数据中的一条数据。

2) 特征处理

流量的识别是通过数据流量中的行为特征判断的,需要将与 IP 地址与网络端口号相关的特征进行删除处理[24]。因此,本文将 Flow ID、Src IP、Src Port、Dst IP、Timestamp 等网络标识数据特征进行删除。在进行分类时,经计算分析得知数据集中有 10 个值为 0 的特征,对网络数据分类计算没有太大的帮助,将这些特征删除。需要删除的特征如表 1 所示。

表 1　值为 0 的特征

序号	特征	序号	特征
1	Bwd PSH Flags	6	Fwd Pkts/b Avg
2	Fwd URG Flags	7	Fwd Blk Rate Avg
3	Bwd URG Flags	8	Bw d Byts/b Avg
4	CWE Flag Count	9	Bwd Pkts/b Avg
5	Fwd Byts/b Avg	10	Bwd Blk Rate Avg

3) 数据集合并

为了获取具有多样性的流量识别数据集,把异常值处理过的新的数据集进行合并,在合并时为了满足本文提出的模型,将数据集中 14 种类别的攻击数据统一标记为 Abnormal,而仅有的 1 种正常数据标记为 Normal。合并后流量为 15 727 888 条,其中 Normal 标签流量占比 85.04%,Abnormal 标签流量占比 14.96%。

4) 数据集采样

合并后的数据集具有 Normal 标签的流量为 85%,与具有 Abnormal 标签流量比例不平衡。受到计算机性能限制,大量的数据导致建立模型及测试都会耗费大量时间。由于本文提出的方法主要对正常流量进行识别过滤,在训练网络时,不过多地对异常数据进行训练识别,因此选取 3 300 000 条 Normal 标签数据流量与 1 100 000 条 Abnormal 标签数据流量组成具有全新的多样性的小体量数据集。

5) 数据归一化

为了提高模型的稳定性和准确性,在数据集进入模型前将数据集中值进行数据归一化处理,将特征值数据映射到[0,1]内,方便模型计算[25]。归一化公如式(10)所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

式中:$x_{\max}$、$x_{\min}$ 分别为特征值的最大值和最小值;x' 为特征值归一化后的数值。

6) 数据集分割

将归一化后的数据集打乱重新组合,按照 7∶3 的分配比例拆分为训练数据集和测试数据集[26]。将随机种子设置为 1,保证每次迭代训练测试时数据集都是一样的[27]。

3　实验与分析

本文利用 Tensorflow 和 Keras 模型框架建立 IDOA-RBF 神经网络的正常流量过滤模型,使用 CSE-CIC-IDS2018 网络入侵检测数据集并采用十折交叉验证的方法进行验证[28]。

3.1　模型参数设定

本文通过对 IDOA-RBF 正常流量过滤模型多次调整学习率和迭代次数进行实验,得出最优学习率和迭代次数对应的模型性能[29]。如图 4 所示,学习率在 0.1 时准确率最高,达到了 97%。如图 5 所示,在模型迭代 150 次时,模型准确率在相对较短的时间开销内接近最大值,因此设置模型迭代次数为 150 次,模型性能最好。

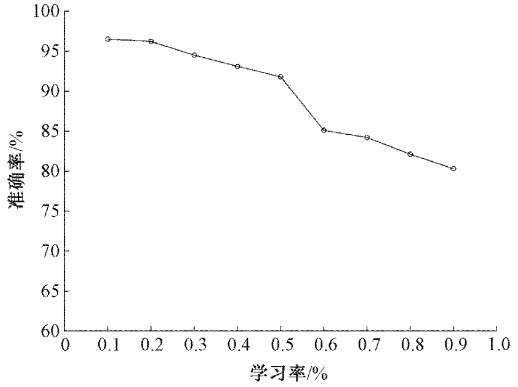

图 4　学习率调参

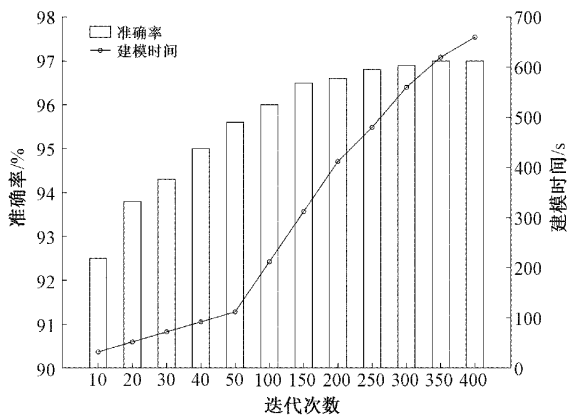

图 5　迭代次数调参

3.2　IDOA 性能分析

为了验证 IDOA 性能,本文使用 Rosenbrock 和 Step 函数为基准,并与 SADOA[30] 和 DOA 进行对比。所有函数的维数都是 30,每个变量都在区间[-100,100]内。Rosenbrock 和 Step 函数分别如式(11)和(12)所示。

$$f(x) = \sum_{i=1}^{n-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \tag{11}$$

$$f(x) = \sum_{i=1}^{n}([x_i + 0.5])^2 \tag{12}$$

将 IDOA、SADOA 和 DOA 分别在 Rosenbrock 和 Step 函数上进行测试。通过 150 次迭代得出 IDOA、SADOA 和 DOA 的性能对比结果。如图 6 所示,在

Rosenbrock 函数上，IDOA 收敛速度明显比 SADOA 和 DOA 快，DOA 在 50 次左右收敛，SADOA 在 30 次左右收敛，而 IDOA 在 10 次左右做到收敛，并且能够找到最优解。如图 7 所示，在 Step 函数上 IDOA 明显比 SADOA 和 DOA 的收敛精度高，而且 IDOA 迭代中快速收敛，在 25 次迭代就已经收敛到最优解。由上所述实验表明，IDOA 在收敛速度和精度上与现有的 SADOA 和原始的 DOA 相比具有更好的性能。

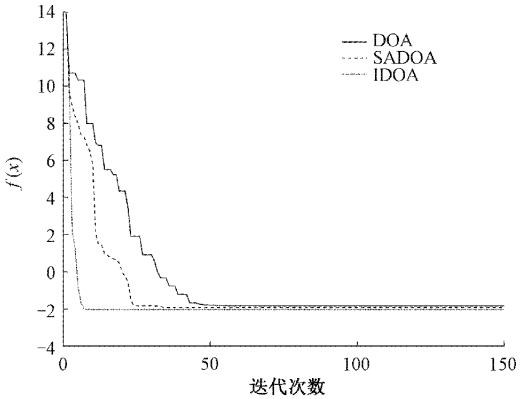

图 6　算法在 Rosenbrock 函数上优化效果

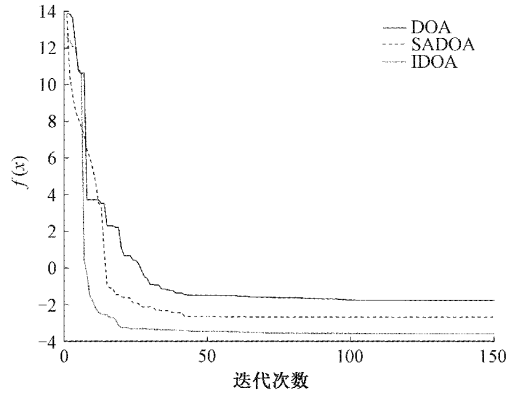

图 7　算法在 Step 函数上优化效果

3.3　模型评价函数

平均绝对误差(mean absolute eror，MAE)作为平均模型性能指标，能够很好的展示出估计值和真实值之间的误差[31]。

均方根误差(root mean square error，RMSE)用来衡量数据值变化的大小程度[32]。

$$MAE=\frac{1}{m}\sum_{i=1}^{m}\left|y_i^k-\bar{y}_i^k\right| \tag{13}$$

$$RMSE=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i^k-\bar{y}_i^k)^2} \tag{14}$$

式中：y_i^k 为样本 k 的第 i 个识别数据值；$\bar{y}_i^k$ 为样本 k 的第 i 个真实数据值；m 为测试集数据数量。

3.4　实验结果分析

为了验证本文提出的 IDOA-RBF 神经网络的正常流量过滤模型的可行性，设置 tanh 函数为模型激活函数，学习率设置为 0.1，最大迭代次数为 150。

1) 性能指标

本文使用准确率、精确率、召回率和 F 值等指标来评价 IDOA-RBF 白流量过滤模型的性能[33]。

2) 实验结果对比

本文将 IDOA-RBF 正常流量过滤模型与现有网络模型进行实验对比分析，得出实验结果。

如表 2 所示，经分析可得，本文提出的 IDOA-RBF 神经网络的正常流量过滤模型与 SADOA-RBF、DOA-RBF、GA-RBF、RBF 和 BP 网络模型相比在整体准确率方面分别提升了 4.1%、6.9%、7.6%、14.0%和 19.8%；在正常流量识别过滤方面分别提高了 6.6%、7.4%，4.2%、7.5%和

表 2　实验对比结果分析

分类器	性能指标	分类器性能	
		Normal	Abnormal
IDOA-RBF	整体准确率/%	93.5	
	建模时间/s	312	
	精确率/%	97.2	89.8
	召回率/%	95.6	85.9
	F 值	96.4	87.8
SADOA-RBF	整体准确率/%	89.8	
	建模时间/s	1 096	
	精确率/%	91.2	87.4
	召回率/%	93.4	66.1
	F 值	92.2	75.3
DOA-RBF	整体准确率/%	87.4	
	建模时间/s	1 328	
	精确率/%	90.5	84.3
	召回率/%	92.4	68.1
	F 值	91.4	75.3
GA-RBF	整体准确率/%	86.9	
	建模时间/s	1 515	
	精确率/%	93.2	80.6
	召回率/%	90.4	70.8
	F 值	91.8	75.4
RBF	整体准确率/%	82.0	
	建模时间/s	2 745	
	精确率/%	90.4	73.6
	召回率/%	64.3	78.8
	F 值	75.1	76.1
BP	整体准确率/%	78.0	
	建模时间/s	3 845	
	精确率/%	85.4	70.6
	召回率/%	70.3	67.9
	F 值	77.1	69.4

13.8%;在建模时间开销方面与其他网络模型相比明显降低了很多。同时 IDOA-RBF 模型可以识别并过滤掉 97.2%的正常流量,占总流量的 72.9%。这得益于 IDOA 的快速收敛能力,能够快速的寻找到全局最优解,构建出拥有高识别精度的正常流量过滤模型。

3) 模型性能分析

如图 8 所示,IDOA-RBF 模型的平均绝对误差比 SADOA-RBF、DOA-RBF、GA-RBF、RBF 和 BP 网络模型分别低 24.1%、25.4%、26.02%、26.42%、27.33%;如图 9 所示,IDOA-RBF 模型的均方根误差比 SADOA-RBF、DOA-RBF、GA-RBF、RBF 和 BP 网络模型分别低 15.2%、15.9%、17.41%、17.75%、19.53%。由上述分析可得,IDOA-RBF 网络模型在各项性能指标均比现有网络模型优良,这得益于 IDOA 具有良好的收敛能力和精度,可以自适应的寻找到更优的模型参数,快速的构建出拥有高识别精度的正常流量过滤模型。由上述表明,本文提出的 IDOA-RBF 正常流量过滤模型可以很好的识别出需要检测流量中的正常流量并过滤掉。

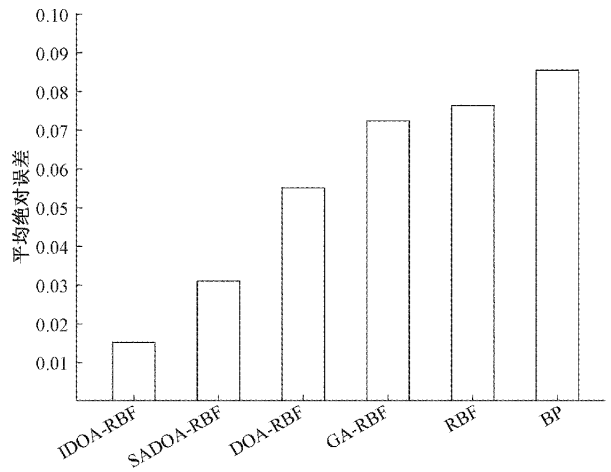

图 8　模型的平均绝对误差对比

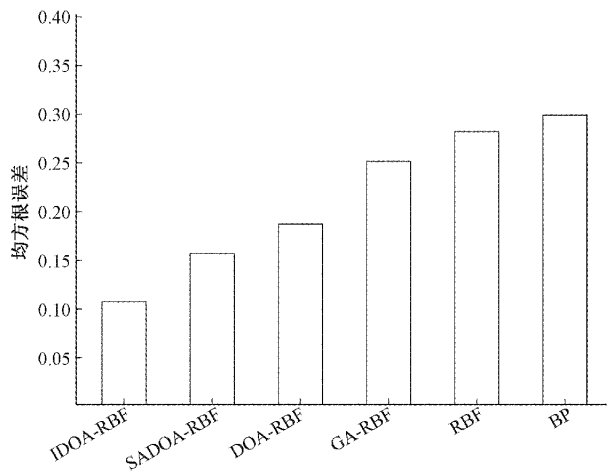

图 9　模型的均方根误差对比

4　结　　论

本文为解决安全检测设备面对大量数据时会出现性能瓶颈的问题。提出一种基于 IDOA-RBF 神经网络的正常流量过滤方法,采用 IDOA 优化 RBF 神经网络的输出权值,提高网络模型的收敛速度和过滤能力。使用测试函数对 IDOA、SADOA 和 DOA 进行实验测试。实验结果表明,IDOA 在收敛速度和精度上具有更好的性能。将构建好的模型方法与 SADOA-RBF、DOA-RBF、GA-RBF、RBF 和 BP 神经网络模型进行对比实验。实验结果表明,IDOA-RBF 模型正常流量识别过滤能力与其他模型相比分别提高 6.6%、7.4%,4.2%、7.5%和 13.8%,而且可以快速地构建出正常流量过滤模型。本文提出的方法减轻安全检测设备的工作负担,在流量安全检测方面具有良好的实际应用价值。目前的网络模型虽然建模时间大大缩短,但是过滤能力提高却不是很明显,在未来的研究中,着重对 RBF 神经网络结构的径向基函数的数据中心 c_i 和扩展常数 σ 的优化,进一步提高模型的过滤能力。

参考文献

[1] 刘海燕,张钰,毕建权,等. 基于分布式及协同式网络入侵检测技术综述[J]. 计算机工程与应用,2018,54(8):1-6.

[2] 董慧. 基于强化学习的网络数据流异常检测数学建模[J]. 电子设计工程,2022,30(4):106-109.

[3] 许学添. 基于模糊约束的网络入侵检测方法[J]. 西安工程大学学报,2016,30(5):627-632.

[4] 赵勇,陈亮,晁萍瑶. 安全存储系统中 TCP 流还原加速策略研究[J]. 西安工程大学学报,2018,32(1):121-125.

[5] 刘博文. 改进后的 BP 神经网络在入侵检测中的应用[J]. 信息与电脑(理论版),2017(14):67-68.

[6] KHAN F A, GUMAEI A, DERHAB A, et al. A novel two-stage deep learning model for efficient network intrusion detection[J]. IEEE Access, 7: 30373-30385.

[7] NAIK B, OBAIDAT M S, NAYAK J, et al. Intelligent secure ecosystem based on metaheuristic and functional link neural network for edge of things[J]. IEEE Transactions on Industrial Informatics, 2020, 16(3): 1947-1956.

[8] 党建武,谭凌. 改进果蝇算法优化加权极限学习机的入侵检测[J]. 系统仿真学报,2021,33(2):331-338.

[9] 杨彦荣,宋荣杰,周兆永. 基于 GAN-PSO-ELM 的网络入侵检测方法[J]. 计算机工程与应用,2020,56(12):66-72.

[10] 柳毅,阴梓然,洪洲. 基于堆稀疏自编码的二叉树集成入侵检测方法[J]. 计算机应用研究,2020,37(5):

1474-1477.
[11] 解男男. 机器学习方法在入侵检测中的应用研究[D]. 长春：吉林大学，2015.
[12] 吴贻淮. 基于神经网络的车载 CAN 网络入侵检测系统的研究[D]. 成都：成都信息工程大学，2018.
[13] KHAN F A，GUMAEI A，DERHAB A，et al. A novel two-stage deep learning model for efficient network intrusion detection[J]. IEEE Access，7：30373-30385.
[14] 钱来，王伟. 一种基于 C-GRU 飞行轨迹预测方法[J]. 电子测量技术，2022，45(10)：87-92.
[15] WANG Y N. Study on neural network optimizing algorithms based on improved wolf swarm algorithms[J]. Journal of Physics：Conference Series，2021，1992(2)：022136.
[16] PERAZA V H，PEÑA D A F，ECHAVARRÍA C G，et al. A bio-inspired method for engineering design optimization inspired by dingoes hunting strategies[J]. Mathematical Problems in Engineering，2021，2021：1-19.
[17] 刘园园，贺兴时. 基于 Tent 混沌映射的改进的萤火虫算法[J]. 纺织高校基础科学学报，2018，31(4)：511-518.
[18] GANG Y. Chaotic butterfly optimization algorithm based on particle swarm optimization[J]. International Core Journal of Engineering，2022，8(3)：420-432.
[19] 张昊，张小雨，张振友，等. 基于深度学习的入侵检测模型综述[J]. 计算机工程与应用，2022，58(6)：17-28.
[20] 王晓峰，董会旭，于岩，等. 基于改进 RBF 神经网络的雷达信号识别[J]. 国外电子测量技术，2022，41(5)：52-56.
[21] RING M，WUNDERLICH S，SCHEURING D，et al. A survey of network-based intrusion detection data sets[J]. Computers & Security，2019，86：147-167.
[22] 张玉清，董颖，柳彩云，等. 深度学习应用于网络空间安全的现状、趋势与展望[J]. 计算机研究与发展，2018，55(6)：1117-1142.
[23] SUN L，ZHOU Y，WANG Y，et al. The effective methods for intrusion detection with limited network attack data：Multi-task learning and oversampling[J]. IEEE Access，2020，8：185384-185398.
[24] 张文哲，张丽娟，陈海倩，等. 基于卷积神经网络的 SSLVPN 流量的识别研究[J]. 电子设计工程，2020，28(12)：144-148.
[25] 王益艳. 基于多方向的各向异性边缘检测算法[J]. 计算机与数字工程，2020，48(1)：167-169.
[26] 司垒，王忠宾，王浩，等. 基于惯性传感组件和 BP 神经网络的防冲钻孔机器人钻具姿态解算[J]. 仪器仪表学报，2022，43(4)：213-223.
[27] CHEN Z M，SUN W J. Research on network intrusion detection based on neural network-based fish swarm optimization algorithm[J]. Telecommunications and Radio Engineering，2020，79(2)：175-182.
[28] 梁本来，朱磊. 基于改进蚁群求解特征子集的入侵检测方法[J]. 计算机应用与软件，2021，38(7)：323-331.
[29] KIM J，KIM H S. Intrusion detection based on spatiotemporal characterization of cyberattacks[J]. Electronics，2020，9(3)：460.
[30] ARAVIND K，MADDIKUNTA P K R. Dingo optimization based cluster based routing in internet of things[J]. Sensors，2022，22(20)：8064.
[31] 鲁思琪，周先春，汪志飞. 改进型自适应全变分图像降噪模型[J]. 电子测量与仪器学报，2022，36(6)：236-243.
[32] 秦永俊，唐增明. 改进的 NetLinX 开放网络动态入侵检测方法[J]. 西安工程大学学报，2017，31(4)：576-581.
[33] 李荣，贺兴时，杨新社. 基于萤火虫算法和高斯扰动的飞蛾优化算法[J]. 纺织高校基础科学学报，2020，33(4)：101-110.

作者简介

钱来，硕士研究生，主要研究方向为无人任务分配、无人机自组网、网络安全。
E-mail：2458619593@qq.com

王伟，副教授，主要研究方向为网络信息安全、网络智能化应用。
E-mail：174456430@qq.com