科学技术与教育 2025年第11期
ISSN: 3079-4455

基于模拟退火优化 CNN 的入侵检测研究

雷田[1]，王提[2]，闫喜海[1]，王珂[1]，闫芮铵[3]
[1]河南工业大学人工智能与大数据学院，河南郑州，中国
[2]解放军信息工程大学电子通信学院，河南郑州，中国
[3]科大讯飞股份有限公司，河南开封，中国

【摘要】入侵检测是信息安全防护的重要环节，卷积神经网络（CNN）可提升检测准确率，但超参数选择是关键难题。本文提出基于模拟退火算法（SA）优化的 CNN 模型，通过 SA 自动调整超参数以快速获取最佳模型。实验表明，该模型在 KDD CUP 99 数据集上的准确率（93.59%）和 F1 值（93.26%）均优于随机搜索和贝叶斯优化算法，验证了其有效性。

【关键词】入侵检测；模拟退火算法；卷积神经网络；超参数优化

1.引言

入侵检测系统（Intrusion Detection System, IDS）概念是 Anderson 首先正式提出的，IDS 主要通过采集网络活动的流量，根据定义好的规则来过滤异常数据，从而实现实时报警[1]。IDS 属于主动防护技术，是网络安全系统中的第二道防线。从检测的手段来看目前的检测技术，主要分为基于异常的检测技术和基于签名的检测技术[2,3]。基于签名的检测技术虽然具有较高的准确率，但是无法检测到未知的入侵攻击类型，而基于异常的检测技术则是可以检测到未知的入侵攻击类型，因此基于异常的检测技术在学术界得到了比较多的关注，越来越多的人投入到了基于异常的检测技术研究中。

与此同时，伴随机器学习与深度学习技术的持续演进，神经网络技术也被逐步引入至入侵检测技术[4,5]——相较于传统机器学习（如 SVM、决策树需人工选特征），神经网络可自动从原始流量中挖掘异常模式（如 TCP 连接的异常时长、数据包大小分布），适配入侵检测中“特征动态变化”的需求。传统入侵检测所采用的神经网络多为浅层结构，在面对大规模网络流量数据时（如 KDD CUP 99 中的海量连接记录），其特征提取能力难以捕捉异常流量的局部模式（如 DDoS 攻击的间歇性特征），不仅导致分类误差率偏高，还因参数规模冗余使模型训练效率较低。为了解决这些问题刘月峰等[6]提出将卷积神经网络（Convolutional Neural Network, CNN）应用到入侵检测中，并取得了较好的结果。在 CNN 模型中主要利用卷积核实现权重的共享，可以有效的加快了模型的训练速度。

CNN 是一种半监督式神经网络，可以有效的检测到未知的入侵攻击类型，但是 CNN 由于网络的复杂性，如何让网络自主选择超参数一直是需要解决的问题[7]，CNN 中的超参数没有固定的形式，每解决一个问题都要搭建不同的模型，对于具有不同数据结构的入侵检测系统，需要搭建多个 CNN 模型来对不同数据进行检测，因此对于如何选择每个模型的超参数是一个巨大的问题，通过将模拟退火算法（Simulated Annealing, SA）优化算法应用到卷积神经网络的超参数选择中可以解决超参数的选择问题，同时也可以提高 CNN 入侵检测模型的准确率。模拟退火算法是一种通用的概率演算法，可以在搜索空间中找到最优解，将 SA 算法利用到卷积神经网络中可以快速得到最佳的网络模型，该模型训练出来的准确率比一般模型较高。

2.相关知识

2.1 卷积神经网络

卷积神经网络（CNN）在图像处理领域始终展现出优异性能[8,9]。与传统的反向传播（Back Propagation, BP）神经网络相比，CNN 能够显著降低网络参数规模，从而更易实现模型优化。典型的 CNN 网络架构包含输入层、卷积层、池化层、全连接层及输出层五个核心模块，各模块在入侵检测数据处理中分工明确：输入层接收经数值化、归一化后的网络流量特征数据；卷积层通过滑窗运算提取流量数据中的局部异常特征；池化层对卷积特征进行降维，保留关键信息的同时减少计算量；全连接层整合全局特征，建立特征与入侵类型的映射关系；输出层输出二分类（正常 / 异常）或多分类（具体攻击类型）结果。这种结构借助感受野机制与权重共享特性，不仅削减了训练参数总

量、压缩训练耗时，还为后续超参数优化降低了搜索空间复杂度。

CNN 的核心优势在于借助感受野机制与权重共享特性，大幅削减模型训练阶段的参数总量，降低参数冗余度的同时避免无效计算，进而压缩训练耗时。但 CNN 的局限性同样突出：其包含大量需人工调试的超参数，且超参数的选择差异会直接影响模型分类精度，需通过专门的优化算法确定适配入侵检测场景的最优超参数组合，常用的超参数优化方法有，随机搜索和贝叶斯优化，利用超参数优化算法可以得到一个最佳模型，将模拟退火算法优化利用到 CNN 模型的超参数选择中，取得了比前两种方法较好的结果。

2.2 模拟退火算法

模拟退火算法来源于固体退火原理，该算法能够快速找到全局最优解，并且能够有效避免局部最优解。模拟退火算法的一般步骤如下：

定义一个函数，初始化一个原始解，将解空间中的参数带入函数从而产生一个新解；

判断原始解与新解的大小；

基于 Metropolis 准则[10]来判断是否接收新解，粒子在温度T时趋于平衡的概率如式(1)：

$$p = exp\left(-\frac{\Delta E}{kT}\right) \quad (1)$$

其中 E 为温度T时的内能，ΔE 为表示内能的改变量，k为 Boltzmann 常数。

当新解被确定为接受时，用新解代替当前解,否则按照 Metropolis 准则来接收新解，然后降低温度T继续下一轮的实验，直到满足终止条件。

2.3 基于模拟退火算法的卷积神经网络模型

在 CNN 网络模型中运用 log 损失函数，其形式如式(2)：

$$J = -\frac{1}{N}\sum_{i=1}^{N}(y_i log\, p_i + (1-y_i)\, log\,(1-p_i)) \quad (2)$$

其中y_i表示第i个数据标签的真实，p_i表示第i个数据标签的预测值。

将上述 log 损失函数 J 设定为 SA 的目标函数，SA-CNN 模型的具体运算流程如下：

将 SA 与 CNN 模型进行结合，其中将J作为模拟退火算法的目标函数，模型的运算步骤如下：

设置 SA 的初始温度T，初始化解J_0

在预设的超参数搜索空间（涵盖卷积层的 Dropout 比例、激活函数类型等关键参数）内搜索并生成新解J_{new}

计算当前解与新解的目标函数差值(3)

$$\Delta f = J_{\text{new}} - J_0 \quad (3)$$

判断Δf是否为负值，如果是负值，就将新解代替为当前解，如若不是，就根据 Metropolis 准则以概率p来接收新解；

若未满足算法终止条件，则降低当前温度阈值，在参数空间重新搜索并生成新解，进入下一轮迭代；若满足终止条件，则输出当前搜索到的 CNN 最优模型。

3.仿真实验与结果

3.1 数据预处理

本实验选用网络入侵检测领域的基准数据集 KDD CUP 99，该数据集每条记录包含 42 个属性，其中 41 个为特征属性，1 个为标签属性。在 41 个特征属性中，9 个属于离散型特征，其余为连续型特征；标签则分为正常数据与 4 种异常攻击类型两类。为验证模型有效性，实验仅采用该数据集 10% 的训练子集与测试子集进行验证，前期需对原始数据进行预处理，具体包括以下三个关键步骤。

数值化处理：

将数据中具有符号特征的数据转化为数值型数据，其中将代表协议类型的属性和连接状态的属性进行 one-hot 编码将数据特征属性转化为具有 53 维的数据。

归一化处理：

对数据进行归一化处理，可将特征取值映射至 [0,1] 区间，有效提升模型学习速率，归一化计算公式如式 (4) 所示：

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

$X_{\max}$ 为属性的最大值，$X_{\min}$ 为属性的最小值，X为属性的原始值，X_{norm} 为归一化后的结果。

PCA 降维：

由于实验所用 CNN 的卷积层基于二维卷积运算，其要求输入数据为四维张量格式，因此需先将预处理后的原始数据重塑为 (None, 7, 7, 1)（其中 None 代表样本数量的动态维度，可适配不同规模的输入数据集）。主成分分析（Principal Component Analysis, PCA）能将高维数据映射至低维空间，且在低维空间中各维度样本方差最大化，可在保留关键特征信息的同时，筛选出对模型分类结果具有显著影响的 49 维数据，满足二维卷积的输入要求。

3.2 网络模块设计

实验设计的 CNN 模型包含 1 个输入层、4 个卷积-池化层、1 个全连接层及 1 个输出层。为缓解模型过拟合问题，在 CNN 网络中引入 Dropout 机制；同时，激活函数类型与卷

积核数量会显著影响模型训练效率与分类精度，因此将上述参数纳入超参数优化范围。因此实验中主要选择对卷积网络 Dropout 比例，激活函数(Relu，ELU)的选择以及 卷积核的个数这些超参数进行筛选，来选择最佳模型。

针对网络流量中 “正常与异常模式的非线性边界”（如正常 HTTP 请求与 SQL 注入请求的特征差异），激活函数需承担 “强化特征区分度” 的核心作用：通过引入非线性映射，将高维流量特征（如 49 维 PCA 降维后特征）转化为可区分的类别空间，避免线性模型无法捕捉“间歇性攻击特征”的缺陷。实验选择 Relu（适配流量特征的稀疏性，加速训练）与 ELU（缓解异常样本少导致的梯度消失）作为候选类型，纳入 SA 超参数优化范围。其中常见的激活函数有修正线性单元（Rectified Linear Unit, Relu）激活函数和指数线性单元（Exponential Linear Unit, ELU）激活函数。

Relu 激活函数的表达式如式(5)：

$$\mathrm{ReLU}(x)=\begin{cases}x & \text{if } x>0\\ 0 & \text{if } x\le 0\end{cases} \tag{5}$$

ELU 激活函数的表达式如式(6)：

$$f(x)=\begin{cases}x, & \text{if } x\ge 0\\ \alpha\left(e^{x}-1\right), & \text{if } x<0\end{cases} \tag{6}$$

针对 CNN 在入侵检测中因 “流量特征维度高（41 维原始特征）、样本类别不平衡（正常样本占比超 80%）” 导致的过拟合问题，实验引入 Dropout 机制：通过随机屏蔽隐藏层中 30%-70% 的神经元（模拟 “流量特征随机缺失” 场景），减少参数冗余并增强模型对 “稀疏异常特征” 的鲁棒性。考虑到 Dropout 比例直接影响异常攻击样本的识别能力（比例过高易漏检、过低易过拟合），将其纳入 SA 算法的超参数搜索空间（搜索范围 0.1-0.6），由算法自动筛选适配 KDD CUP 99 数据集的最优比例。

3.3 评估标准

本实验采用精确率（Precision, P）与 F1 分数（F1-Score, F1）作为模型核心评估指标，结合入侵检测场景中 “异常攻击识别准确率” 的需求，定义指标计算逻辑如下：

$$P=\frac{TP}{TP+FP} \tag{7}$$

$$\mathrm{RE}=\frac{TP}{TP+FN} \tag{8}$$

$$F_1=\left(\frac{\mathrm{RE}^{-1}+\mathrm{P}^{-1}}{2}\right)^{-1}=2\cdot\frac{\mathrm{P}\cdot\mathrm{RE}}{\mathrm{P}+\mathrm{RE}} \tag{9}$$

其中真正例（True Positive, TP）表示将正类预测为正类的个数，假正例（False Positive, FP）表示将负类预测为正类的个数，假负例（False Negative, FN）表示将正类预测为负类的个数。

3.4 实验结果

实验主要以模型的分类准确率为主要评估标准，其中 Rad-CNN 表示基于随机搜索超参数模型，Baye-CNN 表示基于贝叶斯优化算法搜索超参数模型，SA-CNN 表示基于模拟退火优化算法搜索超参数模型，然后利用训练好的模型进行预测，其结果如表 1 所示。

表 1.实验结果

模型	准确率（%）	F1 值（%）
Rad-CNN	91.36	89.21
Baye-CNN	90.08	89.96
SA-CNN	93.59	93.26

从表 1 中可以看出，利用模拟退火算法得出的模型具有较高的准确率和 F1 值，因此可以得知基于模型退火算法选择出来的模型要优于其他两种方法选择出来的模型。

4.总结

利用 CNN 进行入侵检测相对于传统的基于机器学习的入侵检测技术有较高的准确率，但是 CNN 中有许多需要调整的超参数，超参数主要是为了确定模型，因此如何让模型自主选择最优的超参数一直是一个需要解决的问题，传统的超参数选择方法往往不能选择到最好的模型，因此提出一种利用模拟退火优化算法来选择超参数的模型，同时利用 KDD CUP 99 数据集进行仿真实验，结果表明利用 SA 优化算法选出的模型能够选择到更好的模型，从而提高了入侵检测的准确性

参考文献

[1] Anderson J P. Computer Security Threat Monitoring and Surveillance[J]. Fort Washington: James P Anderson Co, 1980.

[2] 李鹏，周文欢.基于 K-means 和决策树的混合入侵检测算法[J].计算机与现代化，2017（12）：12-16.

[3] Zheng K, Cai Z P, Zhang X, Wang Z J, Yang B H. Algorithms to speedup pattern matching for network intrusion detection systems[J]. Computer Communications, 2015, 62(C): 47-58.

[4] 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（06）：93-98.

[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（01）：82-87.
[6] 刘月峰，王成，张亚斌，苑江浩.用于网络入侵检测的多尺度卷积 CNN 模型[J].计算机工程与应用，2019，55（3）：90-153.
[7] 邓帅.基于改进贝叶斯优化算法的 CNN 超参数优化方法[J].计算机应用研究，2019，36（07）：1984-1987.
[8] 李彦冬，郝宗波，雷航.卷积神经网络研究综述[J].计算机应用，2016，36（09）：2508-2515+2565.
[9] 贾凡，孔令智.基于卷积神经网络的入侵检测算法[J].北京理工大学学报，2017，37（12）：1271-1275.
[10] 周爱武，翟增辉，刘慧婷.基于模拟退火算法改进的 BP 神经网络算法[J].微电子学与计算机，2016，33（04）：144-147.