基于改进人工蜂群算法的工业控制系统入侵检测方法

薛迪杰[1] 忽晓伟[1]
XUE Dijie HU Xiaowei

摘 要 工业控制系统入侵检测依赖定义安全规则或签名来识别已知的入侵行为,然而面对新型或未知的攻击时,其检测精度往往受限。为解决这一问题,研究了一种基于改进人工蜂群算法的工业控制系统入侵检测方法。通过工业网络通信技术采集系统网络数据流量,并进行预处理,以确保数据的完整性和准确性。利用改进的人工蜂群算法(IABC)优化搜索路径,从预处理后的数据中同步提取已知或新型的入侵特征。采用循环神经网络模型深度学习入侵特征实现对入侵行为的精准检测。一旦检测到入侵,立即触发预警机制,保障了工业控制系统的安全。仿真结果表明,基于改进人工蜂群算法的工业控制系统入侵检测方法具有较高的收敛速度和高效利用计算资源的能力,显著提升了检测精度,为工业控制系统的安全防护提供了创新且实用的解决方案。

关键词 改进人工蜂群算法;工业控制系统;人工蜂群算法;入侵检测方法;工控入侵

doi:10.3969/j.issn.1672-9528.2025.05.044

0 引言

随着信息技术的飞速发展,工业控制系统(ICS)作为现代工业的核心组成部分,其安全性日益成为关注的焦点。近年来,针对 ICS 的入侵攻击事件频发,给工业生产带来了巨大威胁,因此,研究有效的入侵检测方法显得尤为重要。当前,虽然已有多种入侵检测方法被提出,但仍面临诸多挑战,如检测精度不高、收敛速度慢等问题。

在已有的研究中,黎佳[1]提出了基于多分类器集成的 ICS 入侵检测算法,该算法通过集成多个分类器,利用不同分类器的优势进行互补,从而提高了检测的准确性和鲁棒性。但多分类器的训练和集成过程相对复杂,可能导致检测精度较低。张子迎等人[2]则设计了基于 XGBoost-DNN 的入侵检测架构,该架构结合了 XGBoost 和深度神经网络(DNN)的优点,通过 XGBoost 进行特征选择和初步分类,再利用 DNN 进行深度学习和优化。但 DNN 的训练过程通常较为耗时,特别是在处理大规模数据集时,检测精度可能受到限制。刘胜全等人[3]研究聚焦于深度强化学习,通过利用深度强化学习算法模拟攻击场景,通过训练模型来学习攻击行为和防御策略,从而实现对入侵行为的快速检测和响应。但深度强化学习的训练过程需要大量的时间和计算资源,导致检测精度较低。曹春明等人[4]提出了基于 VAE 和 DLIESN 的入侵检测方法,利用变分自编码器和深度局部异常检测网络进行特征提取和异常检测,但在实际应用中,VAE 和 DLIESN 之间的参数调优和协同工作也需要进一步研究和优化,容易形成低检测精度的局面。

针对上述文献在收敛速度方面存在的缺陷,本文提出了一种基于改进人工蜂群算法的工业控制系统入侵检测方法。该方法通过优化人工蜂群算法的搜索策略和参数设置,实现了对 ICS 入侵行为的快速准确检测。

1 工业控制系统数据预处理

工业控制系统含大量网络装置,安全至关重要。为确保系统安全,准确完整的数据是进行有效入侵检测前提。为了获取这些数据,本研究用 Python 库 packet_tool 收集网络数据流,但数据可能有缺失、噪声和量纲不一致等问题,严重影响分析,干扰检测准确性[5]。因此,构建模型前需预处理数据,从而确保数据完整准确。

为确保数据的相关性和针对性,本研究根据具体的业务需求,精准地去除那些与分析目标无关的数据。对于数据集中出现的缺失值,利用相邻数据点的线性关系,通过计算得出缺失值,提高数据的完整性和准确性,其公式为:

$$y_{\mathrm{interp},i} = y_{i-1} + \frac{(x_i - x_{i-1})}{(x_{i+1} - x_{i-1})} \cdot (y_{i+1} - y_{i-1}) \quad (1)$$

式中:x_i 和 y_i 分别是数据点的横坐标和纵坐标;i-1 和 i+1 分

1. 郑州西亚斯学院电信与智能制造学院 河南郑州 451150
[基金项目] 河南省科技攻关项目(232102220026、232102220071);河南省高等学校重点科研项目(24B120008);河南省教育厅第九批河南省重点学科(检测技术与自动化装置)建设项目(教高〔2018〕119 号)

别是缺失点前后的已知数据点。

为消除不同量纲对数据的影响，需要进一步进行数据标准化处理。采用最大最小值归一化法，将数据转换到同一尺度上，使得不同量纲的数据可以进行比较和分析：

$$y_{\text{norm},i}=\frac{y_{\text{interp},i}-y_{\min}}{y_{\max}-y_{\min}} \tag{2}$$

式中：$y_{\text{norm},i}$ 是第 i 个数据点归一化后的结果；$y_{\min}$ 和 $y_{\max}$ 分别是数据集中的最小值和最大值。通过计算数据的最大值和最小值，将数据映射到 [0,1] 区间内，从而实现数据的无量纲化。

经上述预处理流程，可以获得高质量的数据集，为构建准确的入侵检测模型提供坚实的基础。

2 利用改进人工蜂群算法提取入侵特征

面对日益复杂的网络环境和不断变化的入侵手段，需要采取有效的方法来提取入侵特征，以便及时准确地检测入侵行为，保障系统安全。尽管已经通过完成预处理获得了高质量数据集，但数据量仍然庞大，且入侵特征可能隐藏在复杂的数据关系中。传统方法难以挖掘有效特征。人工蜂群算法（ABC）虽具有自组织、自适应及鲁棒性强等特点，但易陷入局部最优。为此，对 ABC 算法进行改进，优化搜索路径并引入新机制，跳出局部最优，探索更广泛解空间，提高找到全局最优解的可能性。改进算法能从预处理数据中同步提取已知或新型入侵特征，为入侵检测模型提供支持。

改进的人工蜂群算法（IABC）设计融入自组织性与任务分工机制两个核心构造原理。自组织性基于正反馈循环、负反馈调节、系统波动性和多元互动机制四大基本特性，驱动算法自适应与进化。任务分工机制采用并行处理和专门化任务分配，相比传统顺序执行策略，显著提升算法执行效率。改进人工蜂群算法提取入侵特征过程如图 1 所示。

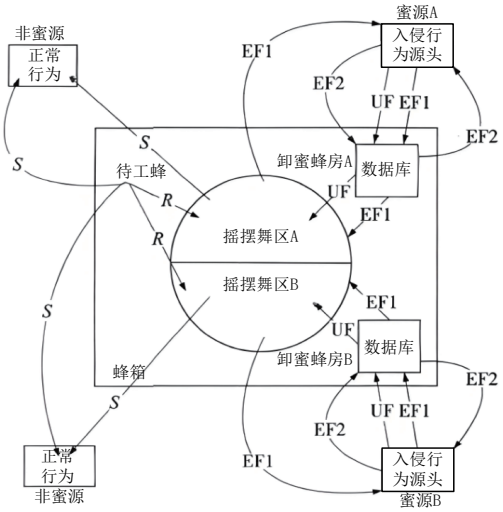

图 1 改进人工蜂群算法提取入侵特征过程

在图 1 所示的场景中，构建一个输出阵列模型，表征网络入侵检测的特征属性：

$$\boldsymbol{O}=\begin{bmatrix} y_{\text{norm},i}(t_1) & f_1 \\ y_{\text{norm},i}(t_2) & f_2 \\ \vdots & \vdots \\ y_{\text{norm},i}(t_N) & f_N \end{bmatrix} \tag{3}$$

在构建的模型中，f_i 表示第 i 个数据点的联合自相关特征。为评估随机生成潜在入侵解的有效性，引入直接封装与间接封装策略，结合 boost 适度评价技术，对解中各入侵信息条目细致评估 [6]。

为优化搜索路径，深入分析种群中表现最优的 YN 只蜂群个体位置信息。蜂群算法模拟蜜蜂觅食行为搜索最优解，个体依据当前位置、邻居位置及适应度信息更新位置。位置的更新规则涵盖了随机搜索、局部搜索和全局搜索等策略。具体表示为：

$$\boldsymbol{O}_i^{(t+1)}=\boldsymbol{O}_i^{(t)}+\alpha\cdot(\boldsymbol{O}_{\text{best}}^{(t)}-\boldsymbol{O}_i^{(t)})+\beta\cdot(\boldsymbol{O}_{\text{neighbor}}^{(t)}-\boldsymbol{O}_i^{(t)})+\gamma\cdot\boldsymbol{r} \tag{4}$$

式中：α、β、γ 是权重系数；$\boldsymbol{O}_{\text{best}}^{(t)}$ 是当前最优个体的位置；$\boldsymbol{O}_{\text{neighbor}}^{(t)}$ 是邻居个体的位置；$\boldsymbol{r}$ 是随机向量。

在搜索过程中，算法会持续评估每个个体的适应度，并据此提取网络入侵特征，可表示为：

$$F=\phi(\{\boldsymbol{O}_i^{(t+1)}\},f) \tag{5}$$

式中：F 是提取的特征集合；$\phi()$ 是特征选择函数；$\{\boldsymbol{O}_i^{(t+1)}\}$ 是当前种群的位置集合；f 是适应度函数。算法设定了最大迭代次数 T，或当最优解的变化小于某个预设阈值时自动终止。算法终止后，会输出最优个体的位置，作为最终提取的网络入侵特征。这些特征能够准确反映入侵行为的特征和模式，为构建高效的入侵检测模型提供有力支持。

3 RNN 模型深度学习入侵特征实现入侵行为检测

工业控制系统中，不同的入侵行为可能在不同的时间点表现出不同的特征，且这些特征之间可能存在相互依赖和关联。传统的入侵检测方法难以捕捉到入侵特征的动态变化和长期依赖关系。循环神经网络（RNN）具有记忆能力，能够处理具有时序信息的数据，因此适合处理这种复杂的入侵特征 [7]。通过深度学习入侵特征，RNN 模型能够更全面、深入地理解入侵行为的模式和特点，从而提高对入侵行为的识别能力，减少误报和漏报的情况，实现更精准的检测。

RNN 模型能够捕捉并利用时间序列中的上下文信息。在 t 时刻，RNN 入侵检测模型的状态可表示为：

$$s_t=\tanh(\boldsymbol{U}F+\boldsymbol{W}s_{t-1}+b) \tag{6}$$

式中：s_t 表示 t 时刻循环单元（隐藏层）的输出状态；tanh

是激活函数，用于引入非线性特性；$\boldsymbol{U}$ 是输入层到隐藏层的权重矩阵，负责将当前输入特征映射到隐藏层空间；$\boldsymbol{W}$ 是隐藏层（环形连接）至隐藏层的加权矩阵，用于捕捉时间序列中的依赖关系；s_{t-1} 是前一时间隐含层的输出；b 是隐藏层的偏置项。

输出层则负责将隐藏层的输出转化为最终的检测结果 O_t。这一转换过程可以表示为：

$$O_t = \sigma(\boldsymbol{V}s_t + c) \tag{7}$$

式中：$\sigma()$ 是 Softmax 函数；$\boldsymbol{V}$ 是隐藏层到输出层的权重矩阵；c 是输出层的偏置项。

这一转换过程确保了 RNN 模型能够准确地将学习到的入侵特征映射到输出空间，从而实现对入侵行为的精准检测 [8]。一旦检测到入侵行为，RNN 模型会立即触发预警机制，向系统管理员发出警报，以便及时采取应对措施，确保工业控制系统的安全。

4 实验

4.1 实验准备

为验证工业控制系统入侵检测方法的有效性，以一家采用高度自动化、智能化和网络化生产线控制与管理系统的汽车制造厂为例进行仿真实验，实验具体场景如图 2 所示。

图 2 生产线控制室数据采集与同步系统架构拓扑图

本文测试实验硬件环境以搭载 Windows 10 操作系统、内存 6 GB、配备 Intel i5 - 8500T 2.10 GHz 处理器的计算机作为实验平台。软件环境方面，选用 Matlab R2023a 作为主要仿真软件，同时引入 PyCharm 2023.1 作为辅助开发工具，借助其强大调试功能和智能代码补全技术提高实验代码开发效率与可读性。

4.2 实验数据

在本研究中，选取入侵检测领域广泛应用的 KDD CUP 1999 数据集（KDD99）作为仿真实验数据基础。该数据集分为训练集（514 002 条记录）和测试集（311 029 条记录），每条数据包含 41 个特征维度，涵盖网络连接属性，如协议类型、服务类型、传输字节数等，最后一维为标签属性，用于标识记录是正常行为还是攻击行为。

为开展研究，从 KDD99 数据集中抽取部分数据作为实验数据，具体信息如表 1 所示。

表 1 实验数据信息

类型	攻击类型描述	训练样本数	测试样本数
Normal	正常网络行为	97 725	52 622
Probe	探测攻击，尝试获取网络拓扑信息	5 566	2 997
DoS	拒绝服务攻击，通过大量请求耗尽网络资源	182 126	98 068
U2R	用户受到根攻击，利用系统漏洞提升权限	76	41
R2L	远程到本地攻击，通过网络入侵获取本地访问权限	2 246	1 210
Unknow	未知类型攻击或异常行为	2 345	1 525

通过选用 KDD99 数据集并预处理，为工业控制系统入侵检测仿真实验构建全面的数据基础。

4.3 实验指标

以检测精度为实验评估指标，从以下两方面对本文方法进行评估：

（1）收敛速度

在工业控制系统入侵检测中，收敛速度快意味着检测系统能更快适应新网络环境或攻击模式，及时提供安全防护。实际应用中，可通过观察算法在训练集上的表现评估收敛速度，收敛越快，算法学习效率越高，检测系统对新环境适应能力越强。

（2）入侵后网络流量偏离正常范围幅值

该指标指入侵行为导致网络流量与正常流量的差异程度，能反映检测方法对这种影响的敏感性和准确性。偏离幅值越大，检测系统越能明显感知入侵行为，检测准确性越高。

4.4 实验结果与分析

在本次工业控制系统入侵检测方法的实验中，对比了 3 种不同算法的检测精度，分别是基于多分类器集成的 ICS 入侵检测算法（方法 1）、基于 XGBoost - DNN 的工业控制系统入侵检测架构（方法 2）以及本文方法。

从图 3 的收敛速度分析可知，方法 1 时间复杂度随迭代次数增加而持续上升，收敛慢且计算成本高；方法 2 虽时间复杂度下降，但整体收敛不理想。本文基于改进人工蜂群算法的入侵检测方法，初期时间复杂度上升后迅速收敛，整体远低于其他两种方法。这表明，改进的人工蜂群算法使本文

方法能智能提取和优化特征，更高效利用资源，实现更快检测速度和更高精度。

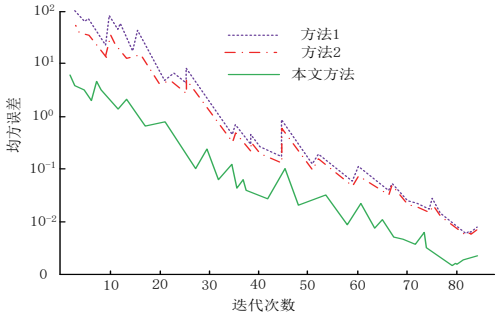

图 3 收敛速度

图 4~6 为入侵攻击后，不同方法检测结果幅值偏离正常范围的情况。

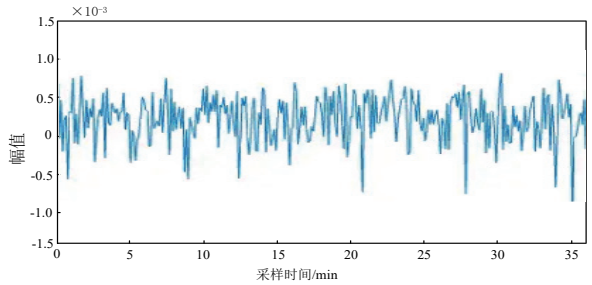

图 4 本文方法检测结果

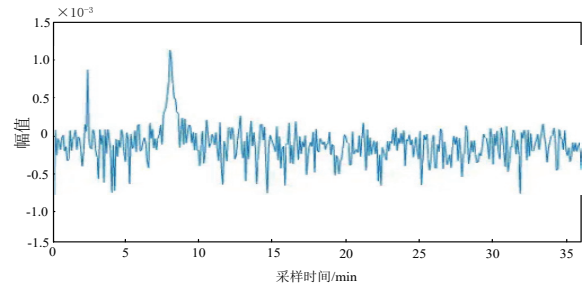

图 5 方法 1 检测结果

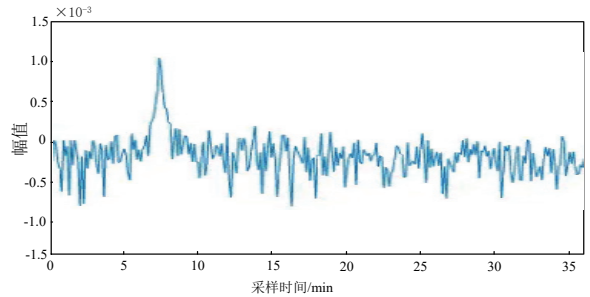

图 6 方法 2 检测结果

分析图 4~6 可知，方法 1 和方法 2 虽能够检测入侵，但偏离范围大，易误报或漏报，影响系统安全。本文基于改进人工蜂群算法的入侵检测方法性能优异，将偏离范围控至最小，优化算法结构精准捕捉特征，提高准确性，降低风险，保障系统安全。引入该算法解决了传统方法局限，提升了性能。综上，本文方法优势显著，为工业控制系统安全提供有力保障。

5 结语

本研究提出了一种创新的工业控制系统入侵检测方法，该方法基于改进的人工蜂群算法，实现了算法层面的深度优化，显著提升了检测效率和准确性。然而，研究也揭示了一些待改进之处，特别是在处理大规模、高维度数据时，算法的运算效率和检测性能可能面临挑战，其在复杂环境下的适应性仍需加强。未来，计划进一步深化对改进人工蜂群算法的探索，从而增强算法在复杂场景下的适应性和运算效率。

参考文献：

[1] 黎佳 . 基于多分类器集成的 ICS 入侵检测算法 [J]. 控制工程 , 2023, 30(6):1105-1111.

[2] 张子迎 , 陈玉炜 , 王宇华 . 基于 XGBoost-DNN 的工业控制系统入侵检测架构 [J]. 哈尔滨工程大学学报 , 2024, 45(11): 2243-2249.

[3] 刘胜全 , 刘博 . 基于深度强化学习的工业网络入侵检测研究 [J]. 东北师大学报 (自然科学版),2024,56(1):80-86.

[4] 曹春明 , 何戡 , 宗学军 , 等 . 基于 VAE 和 DLIESN 的工控系统入侵检测方法 [J]. 计算机工程与设计 , 2023, 44(11): 3283-3289.

[5] 田小芳 . 基于人工蜂群算法的计算机网络 DDoS 攻击检测方法 [J]. 计算机测量与控制 ,2023,31(12):28-33.

[6] 黄兆军 , 曾明如 . 基于 RVM 联合 GSA-SVM 的 ICS 分层入侵检测算法 [J]. 控制工程 ,2022,29(7):1323-1329.

[7] 王华忠 , 田子蕾 . 基于改进 CGAN 算法的工控系统入侵检测方法 [J]. 信息网络安全 ,2023,23(1):36-43.

[8] 石乐义 , 侯会文 , 徐兴华 , 等 . 基于特征选择和时间卷积网络的工业控制系统入侵检测 [J]. 工程科学与技术 , 2022, 54(6): 238-247.

【作者简介】

薛迪杰（1987—），男，河南焦作人，硕士，副教授，研究方向：工业控制系统。

忽晓伟 (1980—)，男，河南南阳人，硕士，副教授，研究方向：机器视觉。

（收稿日期：2024-12-20 修回日期：2025-05-13）