引用格式：姜来为, 赵志天, 杨宏宇. 面向物联网受限终端的机器学习入侵检测方法综述 [J]. 电子科技大学学报, 2026, 55(3): 411-425.
JIANG L W, ZHAO Z T, YANG H Y. A survey of machine learning intrusion detection methods for internet of things restricted terminals[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(3): 411-425.

面向物联网受限终端的机器学习入侵检测方法综述

姜来为[1,2]，赵志天[1]，杨宏宇[1*]

（1. 中国民航大学安全科学与工程学院，天津 300300；2. 中国民航大学计算机与人工智能学院，天津 300300）

摘要：针对物联网（internet of things, IoT）分布式架构与终端资源受限特性导致的脆弱性问题，以及现有综述未系统剖析资源受限场景下物联网终端入侵检测核心瓶颈的缺陷，对终端资源受限 IoT 环境中基于机器学习的入侵检测方法开展研究。首先，解析 IoT 3 层架构，分析 IoT 终端资源受限原因并明确标注数据稀缺与类不平衡、终端计算能力不足、存储资源匮乏等核心挑战；其次，系统梳理近 5 年技术进展，综述了类别均衡与半监督/无监督学习如何缓解标注样本稀缺问题、模型轻量化设计与训练优化算法在降低算力需求方面的突破、数据降维及冗余特征去除技术在内存优化上的有效性，并对比各类方法的优劣；最后，提出构建真实 IoT 专用数据集、处理类间重叠问题等未来方向，为该领域技术深化与工程落地提供参考。

关键词：物联网；入侵检测；机器学习；安全防护
中图分类号：TP393 **文献标志码：**A DOI: 10.12178/1001-0548.2025226

A survey of machine learning intrusion detection methods for internet of things restricted terminals

JIANG Laiwei[1,2], ZHAO Zhitian[1], and YANG Hongyu[1*]

(1. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. School of Computer Science and Artificial Intelligence, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In response to the vulnerability issue arising from the distributed architecture and resource-constrained nature of internet of things (IoT) terminals, and the defects of the core bottlenecks of intrusion detection of IoT terminals in the resource-constrained scenarios that have not been systematically analyzed in the existing reviews. Firstly, the IoT three-tier architecture is analyzed to elucidate the causes of terminal resource constraints, explicitly identifying core challenges such as the scarcity of labeled data and class imbalance, insufficient terminal computing power, and limited storage resources. Secondly, this paper systematically reviews technical advancements over the past five years. It summarizes how class balancing and semi-supervised/unsupervised learning mitigate the scarcity of labeled samples, the breakthroughs of lightweight model design and training optimization algorithms in reducing computational demands, and the effectiveness of data dimensionality reduction and redundant feature removal technologies in memory optimization, while comparing the advantages and disadvantages of various methods. Future research directions, such as constructing realistic IoT-specific datasets and addressing class overlap issues are proposed, providing a reference for technological deepening and engineering implementation in this field.

Key words: internet of things; intrusion detection; machine learning; safety protection

物联网（internet of things, IoT）通过互联网将物理世界中的设备、传感器、物体等连接起来，使其能够相互通信、交换数据并实现智能化操作[1]。从智能家居到工业互联网，IoT 的应用场景日益扩展，极大地促进了社会的进步与经济的增长[2]。然而，IoT 的发展也带来了前所未有的安全挑战：由于 IoT 终端设备通常资源受限且难以部署复杂的安全保护方案，使其成为了网络攻击者的理想目标[3]，

收稿日期：2025−11−11
基金项目：国家自然科学基金民航联合基金重点项目（U2433205）
作者简介：姜来为，博士，主要从事无线自组织网络、网络资源优化和网络空间安全方面的研究。
*通信作者 E-mail: yhyxlx@hotmail.com

这不仅威胁到了个人隐私和数据的安全性，还可能对关键基础设施造成严重破坏。

当前，对于 IoT 设备的主要安全防范方案包括防火墙、数字加密、入侵检测系统（intrusion detection system, IDS）等。IDS 通过实时监控网络，从而识别潜在的攻击行为。常见的入侵检测方法分为两大类：基于签名和基于异常的检测方法[4]。前者通过匹配已知签名识别入侵行为，签名通常设置为恶意代码或异常网络流量的特征，该方法实现较为简单，对已知攻击的识别效果显著，但需持续更新签名库且无法识别未知攻击；后者则通过分析网络流量构建正常行为模型，检测偏离模型的异常行为以识别潜在的攻击行为，该方法无需频繁更新模型且可以识别未知攻击，因此被广泛运用，但其基线构建依赖海量数据，易受数据质量影响，误报率高。

近年来，随着机器学习（machine learning, ML）的发展，ML 方法也被引入入侵检测领域并显现出自身的优势[5]：相比基于异常中的正常流量模型构建方法，ML 利用海量数据训练出的模型来区分正常活动和恶意活动的独特模式特征[6]，以此提高入侵检测的准确性和实时性，还能有效应对未曾识别过的攻击模式，成为当前入侵检测领域的研究热点。

虽然现有基于 ML 的入侵检测文献较多，但已有综述文章[7-9]多根据使用的算法对基于 ML 的网络入侵检测方法进行分类。本文从 IoT 终端的资源受限问题出发，分析资源受限问题对 IoT 入侵检测的影响并系统性综述近年来解决终端资源受限问题的入侵检测方法，最后对此研究方向的未来发展进行展望。

1 资源约束下物联网终端的安全挑战

IoT 的分布式架构使其成为融合物理世界与数字世界的核心载体，但感知层终端设备的硬件资源约束也为安全防护技术的部署带来了根本性障碍。入侵检测作为主动防御的关键技术，其性能直接依赖于计算、存储与数据资源的支撑，而 IoT 终端设备的资源受限特性与传统网络环境存在本质差异，导致传统设备上的方案难以直接适配。

1.1 物联网架构与终端资源受限特性分析

图 1 为 IoT 架构，包含感知层、网络层和应用层。感知层是数据采集终端，由传感器、嵌入式设备等组成，承担原始数据捕获功能，但其硬件成本与能耗限制导致计算、存储资源极度稀缺，是整个架构的核心资源受限节点；网络层以网关、路由器为核心，负责数据传输与初步汇聚，具备基础计算能力但资源仍有限，主要承担数据预处理任务；应用层部署于云端服务器，拥有充足的计算、存储与持续能源供给，负责复杂模型训练与全局决策。本文所述资源受限 IoT 并非将所有 IoT 场景等同于资源受限，而是特指感知层终端及部分边缘节点。为避免表述冗余，后文统一使用终端资源受限 IoT 代指感知层终端及部分边缘节点。

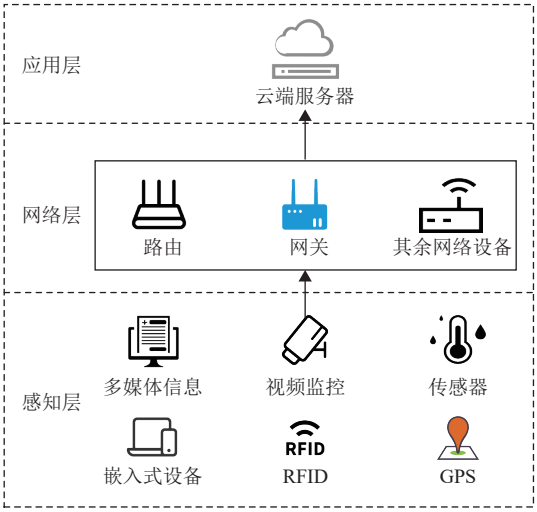


图 1 IoT 架构

为明确 IoT 终端设备的资源约束对基于 ML 的入侵检测的影响，本文通过核心指标对比 IoT 终端与传统网络设备如服务器、PC 机的硬件能力，具体差异如表 1 所示。

表 1 IoT 终端设备与传统网络设备对比

设备	计算能力	内存容量	对入侵检测的影响
传统网络设备	GHz级CPU	GB级	ML模型推理延迟增加10～100倍，复杂模型无法实时运行
IoT终端设备	低主频CPU	KB级	无法存储高维流量数据和参数量大的模型

由表 1 可知，IoT 终端在核心硬件指标上与传统设备存在数量级差距：计算能力不足导致矩阵运算密集的 ML 模型推理耗时过长，无法满足入侵检测的实时性要求；存储资源匮乏使得原始流量数据

在本地存储与高参数量模型的部署难以实现。

1.2　终端资源受限对物联网入侵检测的核心挑战

本文综述内容为基于机器学习的入侵检测方法在 IoT 终端设备场景下的研究进展，即利用机器学习模型从网络流量的行为特征中学习正常与攻击模式，并进行检测与分类的技术路线。基于机器学习的入侵检测不同于基于签名匹配和人工规则判定的传统入侵检测方法，主要依赖数据进行建模并进行分类。因此其性能直接受限于数据、计算和存储资源约束。传统网络设备如服务器的入侵检测方案通常由充足的计算和存储资源支撑，检测精度较高，因此传统网络中基于 ML 的入侵检测方法难以直接部署在终端资源受限的 IoT 环境[10]。具体来讲，基于 ML 的入侵检测在 IoT 场景下面临以下 3 个核心挑战。

1）标注数据资源受限问题：基于 ML 的入侵检测方法虽具备无需频繁更新模型、可识别未知攻击的显著优势，但其正常行为模型的构建与检测性能的稳定高度依赖大量标注数据。IoT 流量数据包含设备状态码、协议字段等专业特征，人工标注需兼具网络安全与 IoT 协议知识，因此 IoT 数据较为珍贵。现有公开数据集中标注完整的 IoT 攻击样本显著少于正常流量[11]，导致 ML 模型因训练数据不足出现过拟合问题。此外，这种严重的标注资源受限问题会导致 ML 模型在训练过程中偏向多数类，最终导致对少数类的漏报率大大提高。因此，需要在 ML 框架下引入针对标注数据不足和类不平衡问题的策略。

2）计算资源受限问题：ML 模型的推理过程涉及大量矩阵运算，而 IoT 终端设备难以支撑复杂模型的计算[12]，对于入侵检测方法而言，这意味着必须在保持高效检测能力的同时，最大限度地减少计算资源的消耗[13]。传统有线网络的入侵检测方法多部署在计算资源较为充足的设备上，并通过复杂计算和大量数据处理方法来提升检测性能[14]。这种资源密集型方法在计算资源受限的 IoT 设备上无法适用。因此，如何在保证高效入侵检测能力的同时，将模型的计算复杂度压缩至终端设备可承受的范围内，是当前研究亟须解决的问题。

3）存储资源受限问题：原始 IoT 网络流量通常包括海量的协议头、载荷和统计特征。直接处理全部的特征不仅会急剧消耗终端设备内存，还会引入大量冗余和噪声特征干扰模型分类性能。终端设备的有限内存，极大地限制了可处理数据的维度和模型规模。这要求入侵检测的过程中采取高效的特征降维和特征压缩能力，在有限的存储空间内保留最具判别力的攻击特征信息。

2　终端资源受限 IoT 入侵检测

本章将从标注数据资源受限、计算资源受限和存储资源受限 3 个核心挑战角度出发，对 IoT 场景下的入侵检测方法开展综述与分析，图 2 为方法分类树。

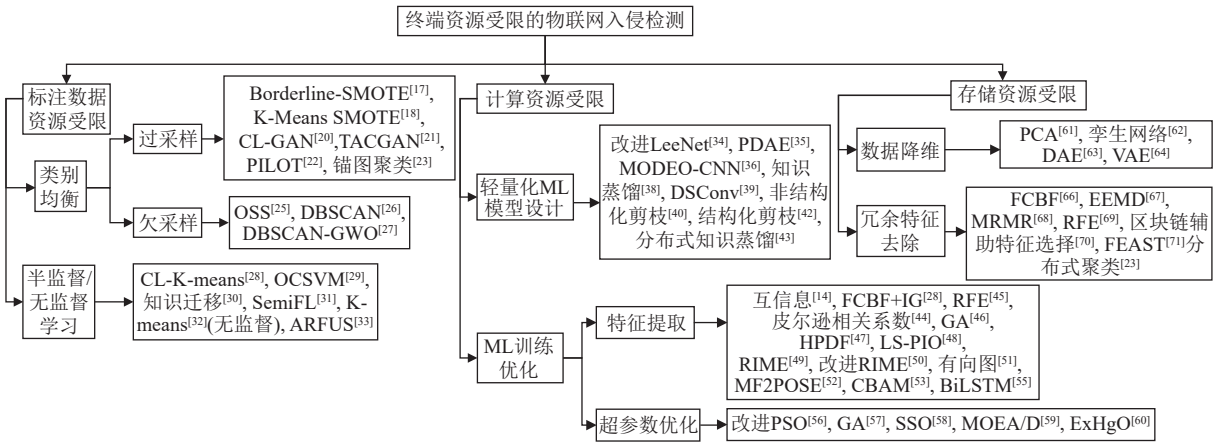


图 2　终端资源受限 IoT 入侵检测分类树

2.1　标注数据资源受限问题及其解决方法

由于 IoT 环境下网络流量数据存在类不平衡问题，即正常流量与异常流量的数量差异显著，类不平衡问题会严重影响入侵检测模型分类性能[15]。标注数据资源受限问题主要关注 ML 模型对少数类攻击的精准检测，解决问题的关键是实现类别均衡或

针对少数类样本进行检测。

2.1.1 类别均衡方法

数据样本的质量与分布均衡性直接决定 ML 模型的训练效果，而标注资源受限的 IoT 场景中，入侵检测数据集普遍存在极端类不平衡问题。为了解决这一问题，类别均衡方法被引入入侵检测领域。类别均衡主要是通过过采样方法对少数类样本进行扩充或使用欠采样方法对多数类样本进行去除优化数据分布。

1）过采样方法

最简单的过采样技术是随机过采样，即随机复制少数类样本来增加少数类样本数量，但随机过采样会由于样本的重复而导致过拟合。合成少数类过采样技术（synth-etic minority over-sampling technique, SMOTE）[16] 是解决过拟合问题的典型方法，其工作原理如图 3 所示。SMOTE 通过合成少数类样本实现数据增强，该方法通过计算少数类样本之间的距离，选择最近的 k 个邻居之一生成新样本。但 SMOTE 会对噪声样本也进行过采样，如果原始少数类样本中包含噪声数据，那么通过 SMOTE 生成的合成样本也会包含更多的噪声，这可能导致模型出现欠拟合的情况。文献 [17] 为了解决上述噪声问题提出了 Borderline -SMOTE 方法，计算数据集中每个少数类样本的 k 个最近邻，根据它们的欧几里得距离划分为安全、危险和噪声 3 类，将在决策边界危险样本的近邻中随机选取一个样本进行插值以增加决策边界区域的样本数量，从而有效提升分类器性能。实验表明，通过 Borderline-SMOTE 扩充的数据集能在不依赖大量负样本的情况下有效地提高 IoT 入侵检测性能，但该方法主要关注靠近决策边界的少数类样本，而忽略了整个数据集的全局分布特性，因此只能解决类间不平衡问题而无法处理类内不平衡问题。文献 [18] 将 K-Means 聚类和 SMOTE 方法结合，先将数据划分为 k 个聚类，把相似数据划分在一个聚类中，计算每个簇的不平衡率，找到安全区域并计算每个安全区域的稀疏性，最后在安全区域中使用 SMOTE 方法进行过采样，该方法可以同时解决类间不平衡和类内不平衡问题，并减少噪声的生成。

除了 SMOTE 方法，生成对抗网络（generative adversarial networks, GAN）[19] 也是常用的过采样方法，其工作原理如图 4 所示。

生成器生成逼真数据试图欺骗判别器，判别器最大化真实数据和生成数据的区分能力，二者交替优化直到生成逼近真实样本的数据。但传统的 GAN 会重复生成相似样本，且生成的数据样本真实性往往达不到预期，无法有效满足数据集质量提升的实际需求。

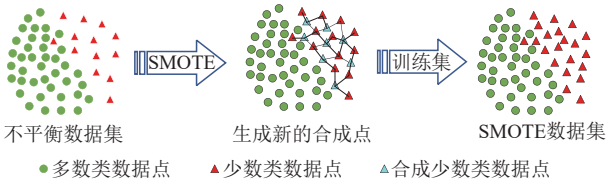


图 3 SMOTE 方法工作原理

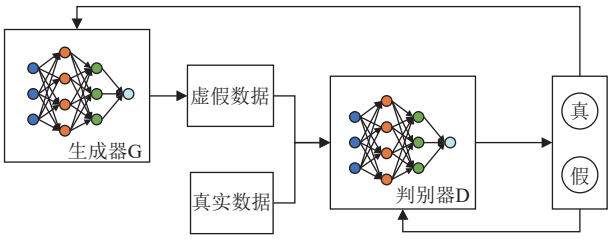


图 4 GAN 实现过采样工作原理

为了解决数据样本真实性问题，文献 [20] 提出了条件生成对抗网络（conditional generative adversarial network, CL-GAN），在 GAN 的基础上引入了 CNN 和 LSTM，构建更具有代表性的数据。为了解决相似性样本问题，文献 [21] 设计了表格辅助分类器生成对抗网络（tabular auxiliary classifier generative adversarial networks, TACGAN），在生成器中引入特征距离损失和特征相似性损失两个损失函数，强制生成数据与真实数据在高层特征空间中保持一致，并在生成器中采用批归一化加速收敛，缓解梯度消失问题。

此外，一些较为新颖的方法也被用于实现过采样。文献 [22] 利用面向漏洞检测的正样本与未标注样本学习模型（positive and unlabeled learning model for vulnerability detection, PILOT）来实现数据增强，通过类间距离原型和渐进式微调为选定的未标注数据生成伪标签来扩充数据集。实验结果表明，该方法相较于几种监督学习方法在 F1 分数上取得了 1.34%～12.46% 的提升。PILOT 方法虽然可以实现数据增强和性能提升，但如果初始分类器预测错误会导致错误标签在后续训练中被反复强化和扩散，进而出现伪标签错误传播风险。因此，PILOT 方法较为依赖初始分类器的分类效果。文献 [23] 引入锚图聚类方法，通过动态锚点机制实现过采样。将原始高维数据映射至低维嵌入空间，随后在

嵌入空间中动态生成锚点矩阵作为虚拟聚类中心，最后通过模糊关系计算新样本与锚点的关联度，从而判定边界样本与未知攻击类别。该方法在只有10% 标注数据的情况下能够实现 82% 的准确率，远远超越传统过采样方法。

总的来说，SMOTE 方法适用于处理中低维度数据，且可以通过 k 值控制生成样本的多样性，但该类方法生成的样本可能加剧类重叠区域的混淆。此外，SMOTE 方法计算开销低，适合部署于 IoT 设备上；GAN 方法则适合处理高维度复杂数据并生成更加逼真的数据，但其计算开销较大，在选择过采样技术时需综合多方面考虑。

2）欠采样方法

欠采样技术通过减少多数类样本的数量，使数据集中不同类别的样本数量趋于平衡。其核心目标是降低多数类对 ML 模型训练的过度影响，提升 ML 模型对少数类的识别能力。随机欠采样是最简单的欠采样技术，即随机去除多数类数据样本实现数据平衡，但此方法因过于随机会丢失部分有价值的信息。为了解决上述问题，基于进化、邻域和聚类的欠采样方法被提出，用于去除多数类样本中重要性较低的数据。

基于进化的欠采样技术通过模拟生物进化机制筛选对分类边界影响最大的样本以保留关键数据分布特征。然而，进化算法需反复迭代优化种群，计算复杂度极高，仅适用于小规模数据集[24]，难以应对 ML 中的大规模数据；基于邻域的方法通过清理局部噪声或边界样本缓解类不平衡，但该类方法只删除了多数类样本中的可疑边界样本，对于 ML 的大规模数据来说，删除边界样本后仍可能存在类不平衡问题[25]。为此，基于聚类的欠采样方法提供了一种更高效的解决方案。该类方法通过聚类算法将多数类样本分组，并用聚类中心或代表性样本替代原始数据。这一过程不仅大幅降低数据规模，还能保留整体分布特性，避免直接删除样本导致的类不平衡残留问题。OSS 算法[25] 是传统的聚类欠采样算法，通过使用 K-means 算法对多数类进行初步聚类，将数据分成多个簇，最后将每个簇中距离簇中心最远的噪声样本移除，实现对多数样本的欠采样。但该方法过度依赖参数调整，不同的 k 值对欠采样的效果影响较大，不恰当的 k 值会导致模型的性能下降。DBSCAN[26] 通过对数据点的密度进行聚类，在每个聚类中基于密度删除样本，这种基于密度的聚类方法可以快速消除噪声及不重要数据，但 DBSCAN 的聚类质量高度依赖手动设置的领域半径，固定参数难以适应不同设备、不同时段流量的动态变化。为了实现动态优化，DBSCAN-GWO[27] 引入灰狼优化器，以轮廓系数最大化为目标为每个数据流窗口实时搜索出最优领域半径。但该方法在 DBSCAN 的基础上添加了额外的元启发式算法，显著增加了计算开销。

针对 IoT 环境下基于 ML 的入侵检测普遍存在的类不平衡问题，基于聚类的欠采样方法通过数据分组与代表性样本提取，实现了数据分布均衡的目标，是未来研究中欠采样技术的首选。常用类别均衡技术总结如表 2 所示。

表 2　类别均衡方法总结

类别	具体方法	优点	缺点
过采样	Borderline-SMOTE[17]	降低边界过拟合	易引入边界噪声，过拟合风险高
	K-Means SMOTE[18]	可保留全局数据分布	聚类计算开销大且依赖聚类质量
	CL-GAN[20]	生成样本质量高	训练不稳定且开销大
	TACGAN[21]	可以解决GAN存在的样本相似性问题	GPU依赖性强
	PILOT[22]	减少80%标注成本	存在噪声伪标签错误传播风险
欠采样	锚图聚类[23]	较少的标注数据实现较高的准确率	计算复杂度显著高于传统过采样
	DBSCAN[26]	可以快速消除噪声	依赖手动设定参数
	DBSCAN-GWO[27]	参数敏感性较低	引入额外模块，增加计算开销

过采样与欠采样是解决类不平衡的两种思路。SMOTE 方法通过合成样本提升少数类表征，但在高维数据下生成质量较差，可搭配数据降维方法生成高质量的少数类样本；聚类欠采样可消除噪声并均衡分布，但对参数过于敏感。GAN 生成数据真实性高，但其依赖 GPU 进行训练。因此，在资源受限的 IoT 终端设备上进行类别均衡优先选择 SMOTE，在具有充足计算资源的条件下优先选择 GAN。

2.1.2　半监督/无监督学习方法

在标注数据稀缺的情况下，半监督学习和无监督学习方法成为应对标注资源受限问题的重要方

法。这两类方法能够利用大量未标注数据，或仅依赖正常流量行为建模，显著降低了对人工标注的依赖，尤其适用于对未知攻击和少数类攻击的检测。

常见的半监督学习方法是自训练，即首先使用标注数据训练一个初始分类器，然后对未标注数据进行预测，将高置信度的预测结果作为伪标注数据加入训练集，迭代优化模型。但自训练方法容易因错误累积从而导致模型性能下降。为了解决这个问题，文献 [28] 选用混合多层模型，第 1 层和第 2 层进行已知攻击的检测分类，第 3 层使用改进的 CL-K-means 模型，先将数据分为不同簇，根据每个簇中哪种类别样本较多打上对应标签，当新的数据出现时根据数据点到簇中心的距离来判断该数据的类别。但基于聚类的 ML 模型依赖簇中心距离计算，分类性能会受到簇分布影响。文献 [29] 将单类支持向量机（one-class support vector machine, OCSVM）和孤立森林结合，使用正常流量训练 OCSVM 模型。该类方法通过集成两个 ML 模型达到较高的检测准确率，但在大规模流量场景下计算复杂度较高，且对于流量自身的噪声敏感。文献 [30] 设计了通过标签驱动的知识迁移，将原始数据复制 N 份，保存一份数据并添加真实标签，其余副本通过循环移位编码生成伪标签数据集；构建教师模型和学生模型，教师模型接受样本和标签的输入，通过自编码器重构特征，学生模型仅接受原始样本学习无标签依赖的特征表示；最后通过均衡加权交叉熵损失函数约束学生模型输出逼近教师模型的判别特征空间，实现检测未知类别攻击的召回率提升 18%。此外，在 IoT 的分布式场景下，半监督学习可以和联邦学习相结合。如文献 [31] 提出了 SemiFL，通过交替训练机制维持伪标签可靠性，每轮聚合后由服务器先用标注数据对全局模型进行微调，再由客户端仅使用接收到的全局模型一次性生成伪标签，并结合置信度阈值筛选与数据增强的一致性训练，使得各终端设备稳定利用未标注数据提升全局性能。但该方法对于置信度阈值敏感，若阈值设置不佳则检查性能也会随之下降。

常见的无监督学习主要通过聚类方法实现，K-means[32] 作为经典的划分式聚类算法，是 IoT 入侵检测中最早被应用的无监督方法之一。其核心逻辑是通过迭代优化将高维 IoT 流量特征划分为预设数量 k 的簇，以簇内样本平方误差和最小化为目标，自动聚合特征相似的流量样本。然而 IoT 流量分布随设备类型、时段动态变化，固定 k 值易导致聚类偏差，而随机初始化中心可能陷入局部最优，且 IoT 流量的高维特征易引发维度灾难导致欧式距离计算的区分度下降。

K-means 等传统聚类方法对距离度量与簇结构高度敏感，且难以适配 IoT 行为随设备类型与时段变化而产生的分布漂移，近年研究逐步转向以深度学习为核心的无监督检测框架。如文献 [33] 提出了 ARGUS，将上下文信息纳入特征学习范围，并采用基于序列自编码器的无监督学习对正常上下文的行为进行重构，从而检测看似正常的隐蔽攻击，同时通过动态阈值策略缓解无监督方法对阈值选择敏感的问题。但该类方法依赖正常行为建模，对伪装成正常行为或与正常行为特征重合的攻击检查性能较差。

半监督学习和无监督学习方法比较如表 3 所示。这两类方法无需依赖大量标注数据，通过学习正常流量的分布特征，识别偏离正常模式的异常流量。但这两类方法存在固有缺陷：若正常流量和攻击流量存在重叠，模型的检测效率会严重下降，使用该类 ML 算法时需要分析所选取的正常流量是否具有代表性。未来标注数据资源受限问题相关研究可以进一步探索半监督与无监督方法的融合，引入主动学习机制优先选择信息量大的样本进行标注，以最大化标注资源的利用效率。

表 3　半监督/无监督学习方法比较

半监督/无监督学习方法	优点	缺点
CL-K-means[28]	无需标注数据	分类性能受参数影响
OCSVM+iForest[29]	提升模型检测性能且实现数据增强	数据增加导致训练时间增加
知识迁移[30]	强大泛化能力	存在标签依赖风险
SemiFL[31]	面向分布式架构	对阈值设置敏感
K-means[32]	计算简单	对参数敏感，易出现维度灾难
ARGUS[33]	对参数不敏感	对正常行为依赖强

2.2　计算资源受限问题及其解决方法

IoT 终端设备普遍存在计算能力弱、能量有限的问题，其处理器主频低、计算核心少，难以承担复杂的模型推理和训练任务。因此，传统的基于 ML

的入侵检测方法往往无法直接部署于 IoT 终端设备，必须从模型、算法两个层面进行轻量化设计与优化。

2.2.1　轻量化 ML 模型设计方法

轻量化模型是以减少参数量、降低计算复杂度和提高推理效率为目的，从而实现资源受限设备上的高效部署。传统的 LeeNet 架构中每个内部模块都以串联的模式运行，这会增加检测时间，文献 [34] 改进了 LeeNet 架构，把 FCNN 层减少到两层并且变成并联的架构，可以有效地减少检测时间。文献 [35] 使用了并行深度自编码器技术，该技术结合了传统自编码器和膨胀自编码器的优点，包含常规卷积滤波器组成的局部特征提取通道和膨胀卷积滤波器组成的周围特征提取通道。这种设计不仅减少了所需参数的数量、内存占用以及处理能力的需求，同时保持了高分类准确度。文献 [36] 设计了 MODEO-CNN 方法，使用整数编码块数量和块内参数的同时采用二进制编码块间连接拓扑，从而避免传统深度神经网络固定层数的冗余结构。这种编码方式允许模型在搜索空间中自适应调整深度与宽度，显著降低内存占用。该方法确保模型在低内存开销的前提下仍然可以保证 F1 分数在 0.97，对内存开销和准确率实现了权衡。

当前最流行的轻量化模型设计方法是知识蒸馏，工作原理如图 5 所示。

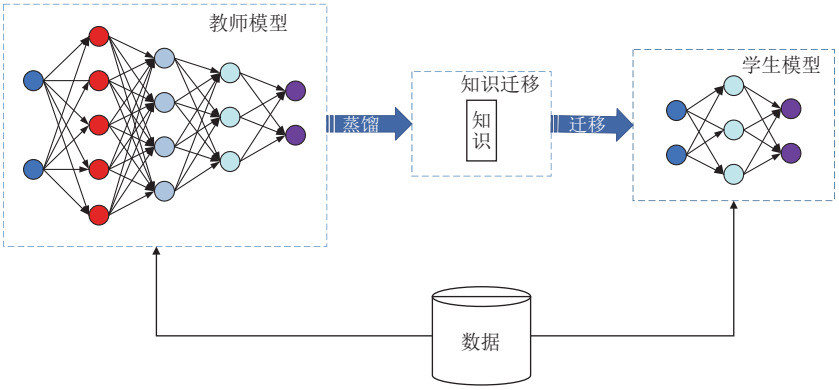


图 5　知识蒸馏工作原理

知识蒸馏是一种教师–学生训练结构，通常是已训练好的教师模型提供知识，学生模型通过蒸馏训练来获取教师的知识，它以轻微的性能损失为代价将复杂教师模型的知识迁移到简单的学生模型中[37]。在 IoT 入侵检测过程中，知识蒸馏可以满足轻量化的需求。文献 [38] 利用一个大规模的 Vision Transformer 作为教师模型，指导一个小规模的学生模型进行训练。在此过程中，学生模型通过学习教师模型的知识来提高自身性能。这种方法使得最终得到的小型模型只需要最少 788 个参数，与知识蒸馏前的大模型相比，减少了大约 90% 的参数数量，同时保持了高检测准确性。文献 [39] 利用大型卷积神经网络学习到的网络流量特征表示，将其转移到深度可分离卷积网络（depthwise separable convolutional network, DSConv）。DSConv 是一种高效的卷积操作，它将标准卷积分解为深度卷积和逐点卷积两个步骤，从而实现参数量减少 96% 的同时计算量减少 95.8%。这种知识迁移过程是通过自我知识蒸馏完成的，即让模型本身担任学生模型和教师模型，无需云端教师模型即可蒸馏。

在使用的模型为深度神经网络时，还可以采用剪枝操作实现模型轻量化设计。模型剪枝通过删除模型中的冗余参数从而有效缩小模型规模，达到显著减少计算和存储需求的目标，并保持模型性能不变。剪枝有两类方法：结构化剪枝和非结构化剪枝。在非结构化剪枝方面，文献 [40] 通过自适应的剪枝方法，在训练过程中根据权重的重要性，动态地增加或删除参数，这种方法能够动态调整模型大小，以减少通信和计算开销，并最小化整体训练时间，同时保持与原始模型相似的精度。但这种方法会带来不规则的稀疏性，导致算法的复杂性增加[41]。在结构化剪枝方面，文献 [42] 通过分析神经网络参数更新的结构性特征来识别并排除不重要的更新，从而显著降低边缘设备与中央服务器间的通信成本。但结构化剪枝会导致剪枝过后的模型准确率下降的问题。

上述方法主要侧重于单个设备内部的模型压缩，未能充分利用 IoT 架构中云端的优势，为将 IoT 的分布式架构利用起来，分布式知识蒸馏成为了当前研究的热点，不同于本地蒸馏，分布式知识

蒸馏将教师模型部署于云端，将学生模型部署于终端设备。如文献 [43] 针对车联网计算资源受限的问题，利用云端部署大语言模型 BERT，提取流量的深层语义特征，并通过软标签远程指导终端设备中轻量级 CNN-BiLSTM 网络进行训练。这种分布式知识蒸馏方式使得 IoT 终端设备仅需极低的浮点运算次数即可获得接近大模型的检测精度，有效解决了复杂模型无法在计算资源受限的 IoT 终端设备上部署的难题。该方法无需考虑终端设备的硬件资源限制，但与云端传递参数会导致额外的通信开销。

从轻量化模型层面优化计算资源的方法总结如表 4 所示。在 IoT 环境下，轻量化的 ML 模型是研究的一个趋势，知识蒸馏凭借较少参数和较高性能成为资源受限场景的最优解。但知识蒸馏方法降低的模型复杂度伴随着检测性能的损失作为代价。未来应探索更高效的知识蒸馏策略，在减少模型参数的前提下，最大程度地减少检测性能的损失。

表 4　轻量化 ML 模型设计方法

方法	优点	缺点
改进LeeNet[34]	改进架构减少检测时间	检测性能下降
PDAE[35]	减少参数数量同时保证高性能	特征融合增加计算开销
MODEO-CNN[36]	低参数量且实时性强	高度不平衡数据集上泛化能力受限
知识蒸馏[38]	大幅减少参数量	依赖云端教师模型
DSConv[39]	无需云端教师模型	自知识蒸馏实时性受限
非结构化剪枝[40]	动态调整模型大小	产生不规则的稀疏权重
结构化剪枝[42]	通过分析减少不必要参数	模型准确率下降
分布式知识蒸馏[43]	无需考虑终端设备硬件瓶颈	引入额外通信开销

2.2.2　ML 训练优化方法

在应对 IoT 终端设备资源受限挑战的解决方案中，轻量化 ML 模型设计虽能显著降低最终部署模型的复杂度，但其性能上限高度依赖于训练阶段所获得的模型参数与特征表示的质量。然而，ML 模型训练过程涉及高维特征处理和大量超参数选择，这与 IoT 终端设备有限的计算资源形成了根本性的矛盾。训练通过特征提取优化减少输入维度、超参数优化提升收敛速度，在训练过程中降低计算开销。

1） 特征提取

特征提取的核心是筛选出和攻击相关的最优特征，降低 ML 模型的计算复杂度，通常以独立指标评估特征重要性或以模型性能为导向通过搜索策略找到最优特征子集。

在特征选择环节常用的独立指标评估方法有信息增益（information gain, IG）法和相关系数法。IG 利用熵来量化特征的不确定性，从而评估特征的重要性，但 IG 主要关注单个特征与目标变量之间的关联性，而忽略了多个特征之间的相互作用。为了解决这个问题，文献 [28] 在 IG 后使用快速过滤的特征选择算法（fast correlation-based filter, FCBF）计算成对特征的对称不确定性，并通过贝叶斯优化高斯过程进行冗余数据处理得到 IG-FCBF 数据集。相关系数法通过量化特征与目标变量之间的线性关系强度，筛选与标签显著相关的特征。文献 [44] 通过计算每个特征与标签的皮尔逊相关系数，按相关系数绝对值降序排列特征，选择前 k 个显著特征用于后续训练，该方法仅需较小的时间复杂度，适用于处理大规模数据集，但该方法易被冗余特征干扰且只能捕捉线性关系。为了解决无法捕捉非线性关系的问题，文献 [14] 通过引入互信息（mutual information, MI）的方法捕捉非线性的特征与标签依赖关系。MI 可以衡量两个变量之间的依赖程度，即通过计算一个变量的信息减少量来表征另一个变量的不确定性。实验验证该方法对非线性的特征与标签之间的依赖关系较为敏感。

总的来说，基于独立指标的特征选择方法仅需根据单独指标评估各特征的重要性以完成筛选，具有较高的计算效率，但该类方法忽略了特征之间的交互作用，筛选出的特征可能对后续模型分类性能产生不良影响。

此外，将模型性能作为评价标准，通过优化策略迭代添加或删除特征、寻找使模型性能最优的特征组合也是特征选择部分常用的方法，且该类方法相较于独立指标评估特征得到的分类性能更优。递归特征消除（recursive feature elimination, RFE）不仅可以实现数据降维，也是较为常用的特征提取方法，文献 [45] 使用 RFE 方法递归地构建模型，并在每次迭代中移除最不重要的特征来识别最佳特征

子集。该方法适用于可以评估每个特征重要性的ML算法，如支持向量机、随机森林等。文献[46]通过遗传算法（genetic algorithm, GA）进行特征选择。GA通过模拟“适者生存”的过程，在解空间中逐步搜索最优或近似最优解。该方法能够在大规模特征空间中找到最优特征子集，有效减少冗余信息并提高分类器的效率和准确性。但GA算法易陷入局部最优解，为了解决上述问题，文献[47]提出了鹈鹕和蜻蜓混合优化算法（hybird pelican and dragonfly optimization, HPDF），在全局探索阶段引入随机扰动因子以增强鹈鹕算法的搜索随机性，从而有效扩大最优特征子集的搜寻空间；加入了鹈鹕算法后准确率相较于GA算法提升了7%。文献[48]提出了LS-PIO的特征选择方法，通过在鸽群优化算法（pigeon-inspired optimization, PIO）上引入禁忌搜索算法和爬山算法来优化特征集，提高准确性的同时还可以减少训练模型用到的特征。文献[49]提出了霜冰优化算法（rime optimization algorithm, RIME），模拟雾状冰的软淞状和硬淞状生长过程，构建软淞状搜索策略和硬淞状穿刺机制。前者通过环境因子β控制特征子集的随机扰动强度，在特征空间中实现宽范围探索；后者通过维度交叉去除低贡献特征，构建最优特征子集。但RIME方法容易陷入局部最优，还存在种群多样性不足等问题。进一步地，文献[50]提出了改进RIME算法，在RIME算法的基础上引入混沌局部搜索策略和交互机制。混沌局部搜索策略根据随机选择的两个不同个体生成混沌扰动更新特征个体位置；交互机制在软霜搜索策略中引入莱维飞行，通过长距离跳跃避免陷入局部最优。这两个改进显著提升了算法的收敛速度和全局探索能力。文献[51]将数据集构建为有向图，利用图中节点和边的信息生成嵌入向量，作为后续训练模型的特征。文献[52]提出MF2POSE方法。此方法结合了多任务特征融合和伪孪生网络，通过融合不同网络分支的信息自动强调对分类最有帮助的特征。文献[53]在IoT入侵检测中引入注意力机制，注意力机制可以让神经网络更多地关注输入中的相关信息，使神经网络更好地利用输入信息[54]，并通过设计通道注意力模块（channel attention module, CAM）和空间注意力模块（spatial attention module, SAM），将二者组成卷积块注意力模块（convolutional block attention module, CBAM），从复杂流量数据中自动提取关键特征并动态调整特征权重。实验表明该方法可以显著提升异常流量的识别能力。文献[55]将注意力机制和双向稀疏自动编码器BiLSTM结合在一起，注意力机制用来捕捉输入数据中的关键特征，并增强对这些特征的学习和表示能力；使用双向稀疏自动编码器，允许从正向和反向两个方向对数据进行编码和解码，从而更全面地提取数据的时间序列特征，增强模型对网络流量行为模式的学习能力，最终达到提升检测性能的目的。

总的来讲，特征提取是优化入侵检测模型性能的关键环节。上述文中涉及的特征提取方法总结如表5所示。近年来，图神经网络、孪生网络及注意力机制等技术逐渐被引入特征选择的研究，为实现最优特征子集的高效筛选提供了新思路：此类方法通过增强特征关联性建模或聚焦关键特征权重分配，显著提升了特征提取效率。但需要注意的是，新模块与算法的引入往往伴随额外的计算开销，如何权衡该类方法带来的性能提升与新增开销已成为未来特征选择技术发展亟待解决的重要研究问题。

表5　特征提取方法总结

特征选择方法	优点	缺点
互信息[14]	独立于模型，可快速预筛选特征	无法处理复杂非线性关系
FCBF+IG[28]	计算开销极低	忽略特征间交互作用
皮尔逊相关系数[44]	计算速度较快，保证实时性	无法处理复杂非线性关系
RFE[45]	适用于较多ML模型	输出结果受分类模型影响大
GA[46]	全局搜索最优特征子集	训练耗时高
HPDF[47]	高鲁棒性和收敛速度	关键参数需手动设置
LS-PIO[48]	引入禁忌搜索优化局部解	参数敏感
RIME[49]	低参数依赖和高计算效率	易陷入全局最优解、种群复杂度不足
改进RIME[50]	有较强全局探索能力	引入新参数需手动调参
有向图[51]	自动学习流量时空依赖	图构建复杂度高
MF2POSE[52]	多任务融合提升判别力	模型参数量大
CBAM[53]	动态加权关键特征	注意力计算增加延迟
BiLSTM[55]	可提取时间序列特征	参数量巨大

2）超参数优化

学习率、卷积核大小、dropout 等超参数的选择直接影响 ML 模型的收敛速度与性能。传统人工调参效率低且易陷入局部最优。超参数优化算法通过在模型训练过程中调用优化算法找到最优模型性能的超参数，从而实现提升检测性能的目的。文献 [56] 对粒子群优化算法（particle swarm optimization, PSO）进行改进，在标准 PSO 算法的基础上对自我学习因子和社会学习因子的计算进行了优化，以便在迭代初期实现较强的发散性并在迭代后期实现较强的收敛性。改进后的算法被用于优化 CNN-LSTM 融合神经网络的超参数，包括学习率、dropout 层参数、卷积核大小和隐藏层神经元数量。实验结果表明，经过改进的 PSO 算法优化的 CNN-LSTM 模型相比其他对比算法，在准确率、召回率等指标上表现更优。文献 [57] 使用了 GA 算法优化卷积神经网络的 7 个关键参数：冻结层的数量、密集层单元数、激活函数、dropout 率、优化器类型、学习率以及训练周期数，并将其性能与协方差矩阵适应进化策略和蜂群优化算法等进行了对比。最终结果表明，经过 GA 优化的超参数显著提升了模型的各项性能指标，特别是在精确率和召回率方面表现突出。文献 [58] 提出了一种结合松鼠搜索优化（squirrel search optimization, SSO）和 Bengio Nesterov 动量的卷积神经网络（SSO-BNMCNN）方法。该方法特别针对卷积层的滤波器尺寸进行了优化，通过 SSO 动态调整滤波器大小，提高了模型对复杂特征的学习效率。此外，此方法中权重更新采用了基于 Bengio Nesterov 动量的策略，进一步增强模型训练过程中的稳定性和收敛速度。实验表明优化后的模型达到了 0.91 的 F1 分数，精确率为 0.96。文献 [59] 利用改进的多目标进化算法/分解（multi-objective evolutionary algorithm based on decomposition, MOEA/D）自动优化和构建 CNN 模型，将 CNN 参数编码为染色体，通过遗传操作如交叉和变异来探索不同的网络架构，并使用准确率强调机制保留性能更优的个体且不牺牲多样性。文献 [60] 提出扩展饥饿游戏搜索优化算法 ExHgO，根据参数组合的质量动态模拟个体饥饿度，并采用开发权重和探索权重平衡全局探索和局部开发。开发权重聚焦当前最优区域精细搜索，探索权重则促进向新区域跳跃，避免陷入局部最优解。该算法可以显著提升逃离局部最优的能力，仅需 150 轮训练即可实现 99.5% 的检测准确率。

关于超参数优化部分的总结如表 6 所示。综上所述，超参数优化主要是为了优化模型训练过程中的超参数以提高模型的分类性能。但通过算法优化实现性能提高往往会带来较大的计算开销。未来研究应致力于开发自适应优化框架，在保障模型分类性能的前提下通过压缩算法实现轻量化算法部署，进而实现 IoT 环境下轻量化入侵检测。

表 6　超参数优化方法总结

超参数优化方法	优点	缺点
改进PSO[56]	实现简单，参数少	易陷入局部最优
GA[57]	全局搜索能力强	收敛速度慢
改进SSO[58]	跳出局部最优能力突出	计算开销大
MOEA/D[59]	多目标优化效果最佳	算法复杂度高
ExHgO[60]	动态适应多场景	计算开销大

2.3　存储资源受限问题及其解决方法

IoT 终端设备的 KB 级内存容量难以承载基于 ML 的入侵检测方法中高维、海量的网络流量数据。原始网络流量包通常包含大量特征，直接对全部特征进行处理将加剧内存负载，导致模型训练与推理过程难以在终端有效部署。更为严峻的是，高维数据中往往存在大量冗余与无关特征，它们不仅占用存储空间，更会引入噪声干扰，降低模型的泛化能力与检测性能。因此，存储资源受限问题已成为制约 IoT 入侵检测模型实际应用的瓶颈之一。为了突破这一瓶颈，当前研究的核心思路是在尽可能保留关键判别信息的前提下，对原始数据进行压缩与提炼。主要解决方法分为两类：数据降维和冗余特征去除。

2.3.1　数据降维方法

数据降维通过降低数据的维数，实现降低输入数据的存储开销。目前主要有 3 类方法用于数据降维：1）线性降维，代表方法是主成分分析（principal component analysis, PCA）[61]。PCA 通过正交变换将原始特征投影至方差最大的正交方向实现线性维度压缩。其核心优势在于计算过程仅需对协方差矩阵进行特征分解而无需复杂的迭代优化。然而，PCA 只能捕捉线性关系，无法处理 IoT 流量中复杂的非线性交互特征。2）监督式降维方法即孪生网络[62]。孪生网络基于共享权重的双分支架构，通过对比损失函数约束同类样本在嵌入空间的相似性，生成低维判别特征。但孪生网络的双分支结构导致训练复杂度较高，需权衡模型深度与设备计算

能力。3）非线性降维，常用的方法有自编码器（autoencoder, AE）和变分自编码器（variational autoencoder, VAE）。文献 [63] 使用深度自动编码器（deep auto-encoder, DAE）将原始数据从 77 个维度降低到 33 个维度，其自动编码器由编码器和解码器两部分组成，编码器将输入数据转换为低维表示，而解码器则从该低维表示中重建原始输入。编码器–解码器架构可以强制学习低维潜在表示，但 DAE 存在过拟合风险。文献 [64] 通过使用 VAE 实现数据降维，该方法通过卷积和反卷积层组成的对称架构处理输入数据，并将高维数据映射到一个 16 维的潜在空间中，从而实现高效降维。关于数据降维技术的总结如表 7 所示。

总的来讲，数据降维的核心是关键信息保留程度与计算复杂度的权衡。PCA 线性降维效率高，但易造成关键信息丢失；AE 非线性降维能够学习网络流量特征之间的复杂关联，但反卷积层又会加大设备能耗。针对 IoT 终端设备资源受限的特点，后续研究应开发具有动态自适应能力的降维算法。此类算法可根据网络流量特征或设备资源状态，动态调整关键信息保留比例或切换降维方法，在保障检测性能的前提下降低降维过程中的计算复杂度。此外，这类算法可以允许在资源紧张的情况下适当牺牲少量精度以换取显著的计算效率和实时性。

表 7　数据降维方法总结

数据降维方法	优点	缺点
PCA[61]	计算高效且无需调参	仅捕捉线性关系，忽略低方差重要信息
孪生网络[62]	保留关键判别信息	依赖初始样本对
DAE[63]	捕捉复杂非线性关系	导致过拟合问题，且存在梯度爆炸/消失问题
VAE[64]	实现高效降维	概率采样易导致输出细节丢失

2.3.2　冗余特征去除方法

在基于 ML 的入侵检测过程中，数据特征空间中的冗余特征会掩盖真实攻击行为的判别，引入噪声干扰，使模型过度关注无关细节。虽然这些冗余信息保证了模型对数据的充分理解，但过多冗余信息会导致计算资源和存储资源的浪费[65]。因此，在原始数据集上使用去除冗余特征算法来减少输入特征的数量，可以使 ML 算法能够更快地训练，降低模型复杂性并使其更易于解释。

去除数据集中冗余特征所用的方法通常是对高相关性特征进行选择，降低后续模型处理特征的复杂度。由于 IoT 终端设备资源受限，在面对存在大量冗余数据的数据集时，检索性能往往被限制，因此进行冗余特征去除对提升模型性能有着重要作用。文献 [66] 选择使用 FCBF 方法计算特征之间的对称不确定性（symmetrical uncertainty, SU），根据 SU 值去除相关性较高的冗余特征。但该类方法仅考虑成对特征冗余，对非线性依赖冗余特征去除效果弱。同样是减少高相关性特征，文献 [67] 将增强过的经验模态分解（ensemble empirical mode decomposition, EEMD）和主成分分析 PCA 相结合。EEMD 首先将输入特征分解成多个本征模态函数（intrinsic mode functions, IMFs），之后 PCA 对 IMFs 进行处理，保留最具代表性的特征并且移除那些与其他特征高度相关的冗余特征。该方法去除冗余特征效率较高，但存在计算量大和依赖参数的问题。文献 [68] 使用最大相关性最小冗余性（max-relevance and min-redundancy, MRMR）特征选择算法计算最小冗性和最大相关性，在提升模型性能的同时减少不必要的特征，从而简化模型结构并提高计算效率。但该方法需要对所有特征计算最小冗余性和最大相关性，二者都需要通过计算互信息得到，因此计算复杂度较高。文献 [69] 使用 RFE 算法反复训练模型并评估特征的重要性，在每次迭代中逐步剔除贡献度最低的特征，最后筛选出最佳特征子集。该方法适用于可以评估每个特征重要性的 ML 算法如支持向量机 SVM、随机森林等。

上述的冗余特征去除方法主要聚焦于单个终端设备内部的数据处理，分布式特征选择方法是当前的研究热点。分布式特征选择通过多节点协作，从整体层面去除无效特征。如文献 [70] 提出了一种区块链辅助的分布式特征选择算法。该方法允许各 IoT 节点基于本地数据独立评估特征重要性，通过轻量级共识机制对冗余特征进行去除并最终筛选出全局最具区分度的特征子集。这不仅避免了本地模型陷入局部最优，更从源头上大幅减少了全网数据的存储与传输开销。文献 [71] 提出了面向联邦学习的高效特征选择框架 FEAST，引入条件互信息计算特征——标签相关性与特征间冗余，并采用多轮迭代的贪心选择策略逐步筛选全局特征子集。

但该方法仍依赖联邦学习中的多节点交互，因此会引入额外的通信开销。此外，分布式聚类方法也被用于处理海量日志样本带来的存储压力。如文献 [23] 提出了基于锚图的分布式聚类方法，通过在终端设备提取少量具有代表性的锚点来替代海量原始数据进行全局分析，有效解决了终端设备资源受限的问题。

关于冗余特征去除方法对比如表 8 所示。总的来说，去除冗余数据不仅可以加快特征选择的速度，还能减少冗余数据对分类模型性能的影响。但去除冗余特征引入的辅助模块会产生额外计算开销，即便使用分布式特征选择和分布式聚类仍引入了额外的通信开销，最终应根据实际需求进行权衡。

表 8　冗余特征去除方法对比

冗余特征去除方法	优点	缺点
分布式聚类[23]	少量标签也可以工作	引入额外通信开销
FCBF[66]	无需模型训练	忽略特征交互
EEMD[67]	适合去除非线性冗余特征	计算量大，参数调优复杂
MRMR[68]	可以生成低冗余、高判别力的特征子集	需计算所有特征的互信息，复杂度较高
RFE[69]	适用于较多ML模型	输出结果受分类模型影响大
区块链辅助特征选择[70]	多个设备均减少特征冗余	引入额外通信开销
FEAST[71]	特征选取高效	引入额外通信开销

3　研究展望

本文通过对现有国内外文献调研，深入分析了终端资源受限 IoT 场景下基于 ML 的入侵检测方法面临的主要问题以及解决方法，并比较了这些方法的优缺点。尽管目前相关研究已经取得了显著成效，但仍有待进一步探索的研究方向。

在标注数据方面，数据质量和可用性是基于 ML 入侵检测的基础，当前研究仍面临着几个瓶颈问题：首先是缺少 IoT 入侵检测专用的数据集，这导致模型的泛化能力受限，易在实际部署中出现过拟合问题。未来可以构建 IoT 环境专用的真实数据集解决泛化能力受限的问题。其次是现有类别均衡方法容易带来噪声放大和类间重叠问题，进而出现决策边界模糊的问题，未来可以引入以边界为约束的数据生成或过滤方法，使生成样本可以落在非重叠区域。再次，现有研究方法大多针对常规的类不平衡问题，而在实际的 IoT 环境下可能存在极端类不平衡问题，未来研究可以引入小样本学习思想，仅通过极少量的数据进行数据生成和入侵检测。最后，半监督学习中引入的伪标签可能存在伪标签错误传播的风险。未来可采用软标签结合不确定性评估，抑制错误累积。

在计算资源受限方面，尽管知识蒸馏、剪枝、量化等轻量化技术能够有效压缩模型规模并降低推理阶段的计算与能耗开销，但过度压缩往往牺牲特征提取能力，导致检测精度下降。鉴于 IoT 终端设备的硬件限制，轻量化模型是未来发展的趋势，未来可根据攻击风险等级触发分层检测，仅对高风险样本启用复杂模型。此外，资源受限条件下训练出的模型往往对噪声和对抗样本更加敏感，未来可在损失函数中加入鲁棒性和可解释性，实现分类性能、鲁棒性和可解释性的三重保证。

在存储资源受限方面，目前虽然有特征降维和冗余特征去除方法，但这些方法默认特征是静态的，难以适应 IoT 场景下攻击的动态变化。未来可将动态特征选择作为研究方向，根据当前流量分布动态调整特征子集，在保持低存储占用的前提下实时捕捉攻击特征。此外，有限的存储资源易导致模型出现“灾难性遗忘”，未来可以仅保留关键样本或模型参数使模型在有限的存储预算前提下持续识别攻击。

此外，当前基于 ML 的入侵检测方法对于 APT[72] 这类攻击行为特征难以提取、攻击渠道多元化、攻击空间不确定特点的攻击检测效果较差。未来可以重点研究多源异构融合的溯源入侵检测方法，使 IoT 入侵检测在真实场景具备应用价值。

4　结束语

本文围绕资源受限的 IoT 环境下的入侵检测方法中面临的挑战和需要解决的问题进行分析，并重点对标注资源受限、计算和存储资源受限 3 类主要问题的现有研究进行文献综述与分类，并分析各类方法的优势与局限，最后讨论专用数据集构建、隐私保护优化、复杂攻击检测、轻量化与性能平衡等未来重点方向，为推动 IoT 入侵检测深入研究和实际应用提供有价值的参考。

参 考 文 献

[1] 郭佳琦, 马智, 王文胜, 等. 云辅助物联网环境下可验证的安全图像检索[J]. 通信学报, 2025, 46(3): 28-44.
GUO J Q, MA Z, WANG W S, et al. Verifiable secure image retrieval in cloud-assisted IoT environment[J]. Journal on Communications, 2025, 46(3): 28-44.

[2] 王鹏, 宋亚飞, 王晓丹, 等. FATIDS: 面向类不平衡样本的物联网入侵检测方法[J]. 吉林大学学报（工学版）, 2025, 52(12): 3986-3999.
WANG P, SONG Y F, WANG X D, et al. FATIDS: An IoT intrusion detection method for class-imbalanced samples[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 52(12): 3986-3999.

[3] LI W J, ROSENBERG P, GLISBY M, et al. Designing energy-aware collaborative intrusion detection in IoT networks[J]. Journal of Information Security and Applications, 2024, 81: 103713.

[4] 姜来为, 顾海洋, 谢丽霞, 等. 机器学习在 WSN 入侵检测中的应用研究[J]. 西安电子科技大学学报, 2024, 51(4): 206-225.
JIANG L W, GU H Y, XIE L X, et al. Application of machine learning in WSN intrusion detection[J]. Journal of Xidian University, 2024, 51(4): 206-225.

[5] HAZMAN C, GUEZZAZ A, BENKIRANE S, et al. Toward an intrusion detection model for IoT-based smart environments[J] Multimedia Tools and Applications, 2024, 83: 62159-62180.

[6] ZHANG Z Q, ZHENG L Y, TAN H N, et al. GBFKAN: An adaptive multilayer interpretable architecture for intrusion detection in various internet of things scenarios[J]. IEEE Internet of Things Journal, 2025, 14(8): 30379-30397.

[7] KIKISSAGBE B R, ADDA M. Machine learning-based intrusion detection methods in IoT systems: A comprehensive review[J]. Electronics, 2024, 13(18): 3601-3623.

[8] 谭振江, 高萌. 基于深度学习的网络入侵检测系统综述[J]. 吉林师范大学学报（自然科学版）, 2024, 45(4): 104-110.
TAN Z J, GAO M. Survey on deep learning-based network intrusion detection systems[J]. Journal of Jilin Normal University (Natural Science Edition), 2024, 45(4): 104-110.

[9] 冯光升, 蒋舜鹏, 胡先浪, 等. 面向物联网的入侵检测技术研究新进展[J]. 信息网络安全, 2024, 24(2): 167-178.
FENG G S, JIANG S P, HU X L, et al. New advances in intrusion detection technologies for internet of things[J]. Netinfo Security, 2024, 24(2): 167-178.

[10] SINGH A, CHOUHAN P K, AUJIA G S. SecureFlow: Knowledge and data-driven ensemble for intrusion detection and dynamic rule configuration in software-defined IoT environment[J]. Ad Hoc Networks, 2024, 156: 103404.

[11] 潘桐, 陈伟, 吴礼发. 面向不平衡样本的物联网入侵检测方法[J]. 网络与信息安全学报, 2023, 9(1): 130-139.
PAN T, CHEN W, WU L F. IoT intrusion detection method for imbalanced samples[J]. Chinese Journal of Network and Information Security, 2023, 9(1): 130-139.

[12] IMTEAJ A, THAKKER U, WANG S Q, et al. A survey on federated learning for resource-constrained IoT devices[J]. IEEE Internet of Things Journal, 2022, 9(1): 1-24.

[13] KHANDAY S A, FATIMA H, RAKESH N. Implementation of intrusion detection model for DDoS attacks in lightweight IoT networks[J]. Expert Systems with Applications, 2023, 215: 119330.

[14] WANG N, SHI S H, CHEN Y M, et al. FeCo: Boosting intrusion detection capability in IoT networks via contrastive learning[C]//Proceedings of the 41st IEEE International Conference on Computer Communications. [S.l.]: IEEE, 2022: 1409-1418.

[15] 朱诗能, 韩萌, 杨书蓉, 等. 不平衡数据流的集成分类方法综述[J]. 计算机工程与应用, 2025, 61(2): 59-72.
ZHU S N, HAN M, YANG S R, et al. Survey on ensemble classification methods for imbalanced data streams[J]. Computer Engineering and Applications, 2025, 61(2): 59-72.

[16] 石洪波, 陈雨文, 陈鑫. SMOTE 过采样及其改进算法研究综述[J]. 智能系统学报, 2019, 14(6): 1073-1083.
SHI H B, CHEN Y W, CHEN X. Research review on SMOTE oversampling and its improved algorithms[J]. CAAI Transactions on Intelligent Systems, 2019, 14(6): 1073-1083.

[17] ZHANG Y, LIU Q. On IoT intrusion detection based on data augmentation for enhancing learning on unbalanced samples[J]. Future Generation Computer Systems, 2022, 133: 213-227.

[18] DOUZAS G, BACAO F, LAST F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE[J]. Information Sciences, 2018, 465: 1-20.

[19] AUNG A P, WANG X R, YU R S, et al. DO-GAN: A double oracle framework for generative adversarial networks[C]//Proceedings of the 2022 IEEE Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11275-11284.

[20] LI S F, CAO Y, LIU S H, et al. HDA-IDS: A hybrid DoS attacks intrusion detection system for IoT by using semi-supervised CL-GAN[J]. Expert Systems with Applications, 2024, 238(F): 122198.

[21] DING H W, CHEN L Y, DONG L, et al. Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection[J]. Future Generation Computer Systems, 2022, 131: 240-254.

[22] WEN X C, WANG X, GAO C, et al. When less is enough: Positive and unlabeled learning model for vulnerability detection[C]//Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering. Echternach: IEEE, 2023: 345-357.

[23] WU Y X, ZHANG L, YANG L, et al. Intrusion detection for internet of things: An anchor graph clustering approach[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 1965-1980.

[24] LE H L, LANDA-SILVA D, GALAR M, et al. EUSC:

A clustering-based surrogate model to accelerate evolutionary undersampling in imbalanced classification [J]. Applied Soft Computing, 2021, 101: 107033.
[25] FARSHIDVARD A, HOOSHMAND F, MIRHASSANI S A. A novel two-phase clustering-based under-sampling method for imbalanced classification problems[J]. Expert Systems with Applications, 2023, 213(B): 119003.
[26] SCHUBERT E, SANDER J, ESTER M, et al. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN[C]//ACM Special Interest Group on Management of Data. Melbourne: ACM, 2017: 19.
[27] MUSTAFA D H, HUSIEN I M. Adaptive DBSCAN with grey wolf optimizer for Botnet detection[J]. International Journal of Intelligent Engineering and Systems, 2023, 16(4): 409-421.
[28] LI Y, MOUBAYED A, SHAMI A. MTH-IDS: A multitiered hybrid intrusion detection system for internet of vehicles[J]. IEEE Internet of Things Journal, 2022, 9(1): 616-632.
[29] CUI J, XIAO J T, ZHONG H, et al. LH-IDS: Lightweight hybrid intrusion detection system based on differential privacy in VANETs[J]. IEEE Transactions on Mobile Computing, 2024, 23(12): 12195-12210.
[30] YANG Y, CHEN J R, WU R J, et al. TGDCLNet: Teacher-guided denoising contrastive learning network-based IoT network intrusion detection[J]. IEEE Internet of Things Journal, 2025, 12(18): 38509-38525.
[31] DIAO E, DING J, TAROKH V. SemiFL: Semi-supervised federated learning for unlabeled clients with alternate training[C]//Proceedings of the 36th Conference on Neural Information Processing Systems. New Orleans: MIT Press, 2022: 1-14.
[32] 杨俊闯, 赵超. K-Means 聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14.
YANG J C, ZHAO C. Research review on K-Means clustering algorithm[J]. Computer Engineering and Applications, 2019, 55(23): 7-14.
[33] RIEGER P, CHILESE M, MOHAMED R. ARGUS: context-based detection of stealthy IoT infiltration attacks[C]//Proceedings of the 32nd USENIX Security Symposium. California: USENIX Association, 2023: 4301-4313.
[34] SUMAN P, PADHY S, KUMAR N, et al. An improved deep learning-based intrusion detection for reliable communication in VANET[J]. IEEE Transactions on Consumer Electronics, 2024, 71(1): 209-217.
[35] BASATI A, FAGHIH M M. PDAE: Efficient network intrusion detection in IoT using parallel deep auto-encoders[J]. Information Sciences, 2022, 598: 57-74.
[36] LU K D, HUANG J C, ZENG G Q, et al. Multi-objective discrete extremal optimization of variable-length blocks-based CNN by joint NAS and HPO for intrusion detection in IIoT[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(4): 4266-4283.
[37] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. 计算机学报, 2022, 45(3): 624-653.
HUANG Z H, YANG S Z, LIN W, et al. A survey on knowledge distillation[J]. Chinese Journal of Computers, 2022, 45(3): 624-653.
[38] WANG Z D, LI J F, YANG S X, et al. A lightweight IoT intrusion detection model based on improved BERT-of-Theseus[J]. Expert Systems with Applications, 2024, 238(F): 122045.
[39] LI Z Y, YAO W B. A two-stage lightweight approach for intrusion detection in Internet of Things[J]. Expert Systems with Applications, 2024, 257: 124965.
[40] JIANG Y, WANG S Q, VALLS V, et al. Model pruning enables efficient federated learning on edge devices[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 10374-10386.
[41] 赵军辉, 李怀城, 王东明, 等. 物联网中模型剪枝技术: 现状、方法和展望[J]. 物联网学报, 2024, 8(4): 1-13.
ZHAO J H, LI H C, WANG D M, et al. Model pruning in IoT: State-of-the-art, methods and prospects[J]. Chinese Journal on Internet of Things, 2024, 8(4): 1-13.
[42] WU X, YAO X, WANG C L. FedSCR: Structure-based communication reduction for federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(7): 1565-1577.
[43] LI S F, CAO Y, PENG G J, et al. Efficient intrusion detection for in-vehicle networks using knowledge distillation from BERT to CNN-BiLSTM[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 6398-6412.
[44] AZIMJONOV J, KIM T. Designing accurate lightweight intrusion detection systems for IoT networks using fine-tuned linear SVM and feature selectors[J]. Computers & Security, 2024, 137: 103598.
[45] YIN Y, JANG-JACCARD J, XU W, et al. IGRF-RFE: A hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset[J]. Expert Systems with Applications, 2023, 238: 122045.
[46] ASIF S. OSEN-IoT: An optimized stack ensemble network with genetic algorithm for robust intrusion detection in heterogeneous IoT networks[J]. Expert Systems with Applications, 2025, 276: 127183.
[47] SINGH K. Industrial internet of things fortify: Multi-domain feature learning framework with deepdetectnet++ for improved intrusion detection[J]. Computers & Security, 2025, 156: 104506.
[48] ABU A O, ALMOBAIDEN W, SAADEH M, et al. An improved PIO feature selection algorithm for IoT network intrusion detection system based on ensemble learning[J]. Expert Systems with Applications, 2023, 213(A): 118745.
[49] SU H, ZHAO D, HEIDARI A A, et al. RIME: A physics-based optimization[J]. Neurocomputing, 2023, 532: 183-214.
[50] WANG L, XU J L, JIA L Y, et al. Multi-strategy RIME optimization algorithm for feature selection of network intrusion detection[J]. Computers & Security, 2025, 153: 104393.
[51] NGUYEN H, KASHEF R. TS-IDS: Traffic-aware self-supervised learning for IoT network intrusion detection[J]. Knowledge-Based Systems, 2023, 279: 110966.

[52] ZHANG J, CHEN R, ZHANG Y, et al. MF2POSE: Multi-task feature fusion pseudo-Siamese network for intrusion detection using category distance promotion loss[J]. Knowledge-Based Systems, 2024, 283: 111110.

[53] 吴昊, 郝佳佳, 卢云龙. 物联网场景下基于蜜场的分布式网络入侵检测系统研究[J]. 通信学报, 2024, 45(1): 106-118.
WU H, HAO J J, LU Y L. Distributed network intrusion detection system based on honeypot in IoT scenarios[J]. Journal on Communications, 2024, 45(1): 106-118.

[54] 王军, 王华琳, 黄博文, 等. 基于联邦学习和自注意力的工业物联网入侵检测[J]. 吉林大学学报（工学版）, 2023, 53(11): 3229-3237.
WANG J, WANG H L, HUANG B W, et al. Industrial IoT intrusion detection based on federated learning and self-attention[J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(11): 3229-3237.

[55] SWATHI K, HIMA B G. An automated intrusion detection system in IoT system using attention based deep bidirectional sparse auto encoder model[J]. Knowledge-Based Systems, 2024, 305: 112633.

[56] WANG Q, JIANG H Y, REN J D, et al. An intrusion detection algorithm based on joint symmetric uncertainty and hyperparameter optimized fusion neural network[J]. Expert Systems with Applications, 2024, 244: 123014.

[57] LI J, CHEN H, OTHMAN M S, et al. NFIoT-GATE-DTL IDS: Genetic algorithm-tuned ensemble of deep transfer learning for NetFlow-based intrusion detection system for internet of things[J]. Engineering Applications of Artificial Intelligence, 2025, 143: 110046.

[58] WU J, HAIDER S A, YU H, et al. An intelligent IoT intrusion detection system using HeInit-WGAN and SSO-BNMCNN based multivariate feature analysis[J]. Engineering Applications of Artificial Intelligence, 2024, 127(A): 107132.

[59] CHEN Y, LIN Q Z, WEI W H, et al. Intrusion detection using multi-objective evolutionary convolutional neural network for internet of things in Fog computing[J]. Knowledge-Based Systems, 2022, 244: 108505.

[60] JOHNPETER T, KARUPPANAN S. Fuzzy-rule based optimized hybrid deep learning model for network intrusion detection in SDN enabled IoT network[J]. Computers & Security, 2025, 152: 104372.

[61] EDUARDO D L H, EMIRO D L H, ORTIZ A, et al. PCA filtering and probabilistic SOM for network intrusion detection[J]. Neurocomputing, 2015, 164: 71-81.

[62] WANG C H, XU D, LI Z H, et al. Effective intrusion detection in highly imbalanced IoT networks with lightweight S2CGAN-IDS[J]. IEEE Internet of Things Journal, 2024, 11: 15140-15151.

[63] KOU L, DING S S, WU T, et al. An intrusion detection model for drone communication network in SDN environment[J]. Drones, 2022, 6(11): 342-360.

[64] NGUYEN T P, CHO J, KIM D. Semi-supervised intrusion detection system for in-vehicle networks based on variational autoencoder and adversarial reinforcement learning[J]. Knowledge-Based Systems, 2024, 304: 112563.

[65] 储岳中, 汪佳庆, 张学锋, 等. 基于改进深度残差网络的图像分类算法[J]. 电子科技大学学报, 2021, 50(2): 243-248.
CHU Y Z, WANG J Q, ZHANG X F, et al. Image classification algorithm based on improved deep residual network[J]. Journal of University of Electronic Science and Technology of China, 2021, 50(2): 243-248.

[66] KUSHWAHA J P, BHADAURIA S, TAPASWI S. mFCBF based lightweight intrusion detection system for IoT networks[J]. Cluster Computing, 2025, 28: 1-19.

[67] LIU Z Q, MOHIUDDIN G, ZHENG J B, et al. Intrusion detection in wireless sensor network using enhanced empirical based component analysis[J]. Future Generation Computer Systems, 2022, 135: 181-193.

[68] TÜRKOĞLU M, POLAT H, KOCAK C, et al. Recognition of DDoS attacks on SD-VANET based on combination of hyperparameter optimization and feature selection[J]. Expert Systems with Applications, 2022, 203: 117500.

[69] ISMAIL M G, GHANY M A E, SALEM M A, et al. Enhanced recursive feature elimination for IoT intrusion detection systems[C]//2022 International Conference on Microelectronics. Morocco: IEEE, 2022: 193-196.

[70] QIU L, WANG X W, YI B, et al. Towards efficiency and decentralization: A blockchain assisted distributed fuzzy-rough feature selection[J]. IEEE Transactions on Parallel and Distributed Systems, 2025, 36: 1762-1778.

[71] FU R, WU Y C, XU Q Q, et al. FEAST: A communication-efficient federated feature selection framework for relational data[C]//Proceedings of the 2023 ACM International Conference on Management of Data. Washington: ACM, 2023: 1-28.

[72] 陈瑞东, 张小松, 牛伟纳, 等. APT 攻击检测与反制技术体系的研究[J]. 电子科技大学学报, 2019, 48(6): 870-879.
CHEN R D, ZHANG X S, NIU W N, et al. A research on architecture of APT attack detection and countering technology[J]. Journal of University of Electronic Science and Technology of China, 2019, 48(6): 870-879.

责任编辑　刘飞阳