doi:10.19562/j.chinasae.qcgc.2024.11.010

基于证据深度学习的CAN网络入侵检测框架*

石 琴[1,2,3],李志伟[1,2,3],程 腾[1,2,3],张 强[1,2,3,4],王文冲[4]
(1. 安徽省自动驾驶汽车安全技术安徽省重点实验室,合肥 230009; 2. 安徽省智慧交通车路协同工程研究中心,合肥 230000;
3. 合肥工业大学汽车与交通工程学院,合肥 230000; 4. 奇瑞汽车股份有限公司,芜湖 241000)

[摘要] 随着移动通信技术在智能自动驾驶系统中的持续发展,保障车载通信数据的安全已成为交通系统安全的一个重要环节,面对黑客可能通过CAN总线网络远程操控车辆的威胁,现有网络框架虽能检测已知攻击,但在识别未知攻击时表现不佳。为此,本研究提出一种融合证据深度学习的检测框架,该框架由数据预处理模块、数据分析模块和攻击检测模块组成。预处理模块通过独立热编码技术,以提升数据质量和适应性;数据分析模块通过生成对抗网络(GAN)技术增强该框架的泛化能力并模拟攻击场景;攻击检测模块应用了证据深度学习,提高了框架在应对未知攻击时的不确定性处理能力。该框架在开源汽车黑客数据集和基于奇瑞EXEED RX车型自主构建的数据集上进行了测试。实验结果表明,该框架在检测未知攻击时,相比于传统的基于softmax的分类网络综合性能提高了24.5%。

关键词:入侵检测;证据深度学习;不确定度;损失函数

Intrusion Detection Framework for CAN Networks Based on Evidence Deep Learning

Shi Qin[1,2,3], Li Zhiwei[1,2,3], Cheng Teng[1,2,3], Zhang Qiang[1,2,3,4] & Wang Wenchong[4]
1. *Key Laboratory for Automated Vehicle Safety Technology of Anhui Province, Hefei University of Technology, Hefei* 230009;
2. *Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province, Hefei* 230000;
3. *School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei* 230000;
4. *Chery Automobile Co., Ltd., Wuhu* 241000

[**Abstract**] With the continuous development of mobile communication technologies in intelligent autonomous driving systems, securing vehicular communication data has become pivotal for transportation safety. Faced with threats of hackers remotely manipulating vehicles through the CAN bus network, existing frameworks can detect known attacks but falter in identifying location-based attacks. A detection framework integrating evidence-based deep learning is proposed in this paper, comprising data preprocessing, analysis, and attack detection modules. The preprocessing module employs independent hot encoding to enhance data quality and adaptability. The analysis module utilizes Generative Adversarial Networks (GANs) to bolster the framework's generalization and simulate attack scenarios. The attack detection module harnesses evidence-based deep learning to enhance the framework's capability in handling uncertainties from unknown attacks.The framework is tested on an open-source car hacking dataset and a dataset constructed based on the Chery EXEED RX model. The test results show that the framework improves the overall performance by 24.5% in detecting unknown attacks compared to traditional classification probability-based networks.

Keywords: intrusion detection; evidence deep learning; uncertainty; loss function

* 安徽省自然科学基金(2208085MF171)、中央高校基本科研业务费专项基金(JZ2023YQTD0073, PA2023GDSK0112)、安徽省重点研究与开发计划项目(202304A05020087)和北京市自然科学基金(7232222)资助。
原稿收到日期为 2024 年 04 月 30 日,修改稿收到日期为 2024 年 06 月 06 日。
通信作者:程腾,副教授,博士,E-mail:cht616@hfut.edu.cn。

前言

在信息社会背景下,车辆对外界信息交换(V2X)通信和车载网络已成为新兴智能交通系统中的核心组成部分[1]。制造商通过融入诸如控制器区域网络(CAN)等通信协议,为汽车提供了多种智能化技术和服务。但网络连接数量的增加伴随着新型网络攻击手段的频繁出现,给汽车信息安全带来了巨大挑战。

自1985年博世公司开发以来,CAN网络因其简洁、低成本、高效及稳定的特性,成为现代汽车车载网络(IVN)中电子控制单元(ECU)通信的主流协议。CAN网络这种基于广播的消息导向串行通信协议,主要用于处理汽车行为的信息传输。然而,CAN消息的通信架构缺少关于发送者或接收者ECU的消息认证机制,导致其安全性和可靠性受到质疑[2]。

近年来,为应对这些安全隐患,关于CAN网络的入侵检测研究日益增多[3]。许多研究例如文献[4]~文献[6],分别提出基于传统统计学、机器学习和集成学习的CAN网络入侵检测框架。但与基于深度学习的CAN网络入侵检测框架相比,这些传统方法效能略显不足。例如,文献[7]中提出基于时间统计的传统入侵检测框架鲁棒性较差。如文献[8]中所述,消息认证或加密为确保实时响应,可能导致CAN设备过载和显著延迟。文献[9]中提出基于分段联合学习的轻量级方法,尽管可通过远程服务器平衡数据,但使得车辆与服务器之间的计算复杂度高且算力消耗大。

文献[10]表明,基于深度学习的车载CAN网络入侵检测框架具有较大潜力,原因是其能处理大量数据,且不依赖于特定领域的先验知识。然而,根据文献[11]~文献[13],这些入侵检测框架在应对新型或未知攻击时表现不佳,主要因为它们过度依赖训练数据。因此,入侵检测框架在识别数据集中未包含的攻击类型时,可能效率较低。此外,数据集质量下降会影响入侵检测框架性能,主要因为传统基于深度学习的入侵检测框架对噪声和异常数据高度敏感。同时,这些入侵检测框架通常被视为“黑箱”,其决策过程和结果输出缺乏透明性。特别是在需要高准确性和可解释性的应用场景中,透明性尤为重要。此外,如文献[14]所述,这些入侵检测框架在评估未知攻击类型(例如零日攻击)的检测能力时存在不足,揭示了其在提供全面安全防护方面的局限性,因此,本研究着重于研究处理未知攻击类型的车载CAN网络入侵检测框架。

本文提出了一种车载CAN网络开放集入侵检测框架,该框架基于证据深度学习技术,旨在提高对未知或新型攻击的检测能力,并在噪声和异常数据下增强该框架的鲁棒性,以及改善检测决策的可信度评估。本文的主要贡献如下。

(1)本文提出的方法首次将证据深度学习应用在CAN网络入侵检测上,并结合生成对抗网络(GAN)提出了一个鲁棒的CAN网络入侵检测框架,大幅提高了网络检测未知入侵的能力。

(2)本文提出了一种证据不确定度校准的方法,有效缓解了模型过度自信预测的问题,使提出的框架在预测准确时对其预测充满信心,而在预测不准确时表现出不确定。

(3)为全面评估框架的性能,使用两个数据集进行测试:公开的SynCAN数据集和基于奇瑞EXEED RX车型自主构建的数据集。与HyDL-IDS、LDAN、O-DAE和TSP对比分析,结果显示提出的框架在应对车载CAN网络的多种攻击类型(如DOS、RPM、GEAR和Fuzzy)表现出最佳的性能。

1 网络框架

如图1所示,本文提出的入侵检测框架包括3个主要模块:一是数据预处理模块,使用独立热编码技术将CAN ID序列转换为二维图像格式;二是数据分析模块,基于生成对抗网络(GAN)构建,对数据视图进行深度分析并生成相应的重构损失;三是攻击检测模块,基于证据神经网络(ENN)构建,专注于计算分类概率以及不确定度并实施最终的检测决策。具体细节如下。

1.1 数据预处理模块

为提升数据特征的表示质量,并为数据分析模块与攻击检测模块提供更清晰、易于区分的数据,本文中设计了一个基于独立热编码的数据预处理模块。此模块主要将CAN ID序列转换为适合输入到生成对抗网络(GAN)的格式。在此模块中,采用独立热编码技术构建尺寸为N×48的二维图像,如图2所示。在数据转换过程中,将11位的CAN ID转换为一个48位向量,该向量由3个16位向量组成,每个向量独热编码一个十六进制数字。例如,ID:0x268(16进制)被转换为数字2、6、8的3个16位向量。接着,将N(本研究中N设置为48)个48位向量叠加堆叠成一个尺寸为48×48的二维正方形图像。

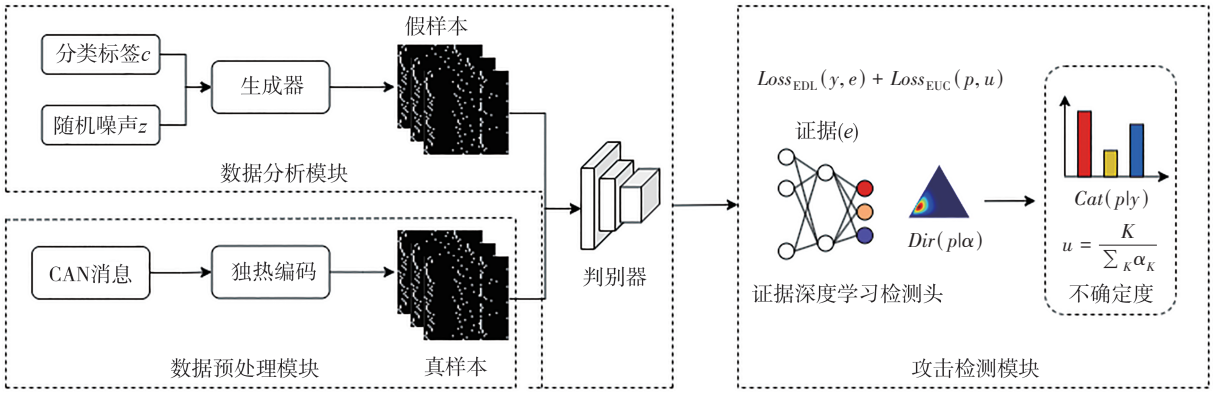

图1　CAN网络入侵检测框架图

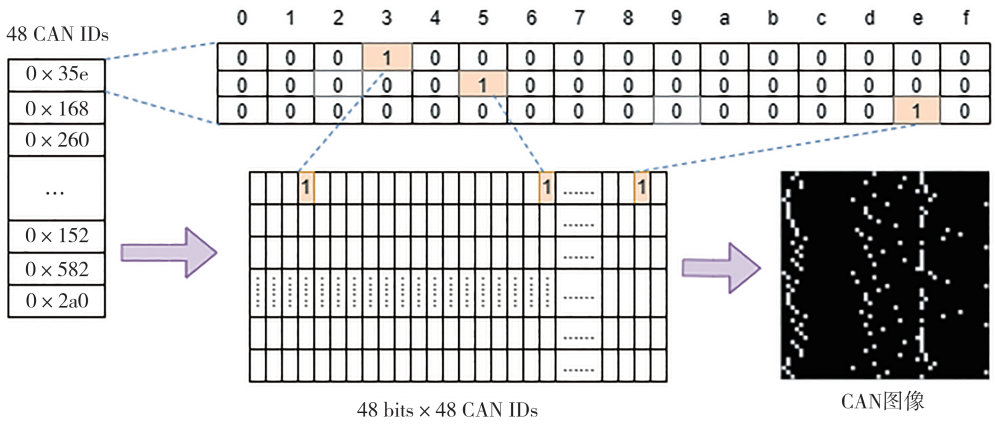

图2　CAN ID独热编码图

该方法在数据差异性的创建上优于传统的直接编码方式,有效区分正常与攻击图像。构成单个CAN消息图像的CAN消息数量是关键超参数,为检测精度与计算成本平衡方面起着决定性作用。为适应图像处理的标准化需求和降低计算负担,本研究将N设置为48。

1.2　数据分析模块

为提升数据质量与有效模拟未知攻击数据,增强框架的泛化性能,本文中设计了基于生成对抗网络(GAN)的数据分析模块(见图1),由生成器和判别器组成。

生成器的目的是捕获数据分布,将来自任意潜在分布(噪声先验)的随机噪声向量z和相应的类标签c作为输入,G使用两者来生成类条件假样本$X_{fake}=G(c,z)$。

本研究设计的判别器比较生成器生成的CAN消息图像与实车数据集样本,从而评估生成图像的真实性,并为攻击检测模块提供决策依据。判别器旨在区分生成器生成的CAN消息图像和数据集中的CAN消息图像。输入为数据集中的"真样本"(X_{real})和自生成器的"假样本"(X_{fake}),输出为两种样本来源$S=\{X_{real}, X_{fake}\}$的概率分布$P(S)$。判别器基于卷积神经网络(CNN)进行特征提取,可附加全连接层,判别器经过训练以最小化L。GAN的损失函数定义为正确来源的对数似然:

$$L=E[\log P(S=\text{real}|X_{real})]+E[\log P(S=\text{fake}|X_{fake})] \tag{1}$$

1.3　攻击检测模块

给定判别器的输出作为输入,位于判别器主干网络之上的证据深度神经网络(ENN)头会预测类别证据,从而制定狄利克雷分布,以便可以确定输入的多类概率和预测不确定性(见图1)。对于开放集推理,高不确定性输入可以被视为未知入侵,而低不确定性输入则通过学习的分类概率进行分类。该模块通过证据深度学习(EDL)损失进行训练,并通过提出的证据不确定性校准方法进行正则化。

当前流行的深度学习模型通常在DNN中使用softmax层来执行分类任务。然而,这些基于softmax

的DNN在分类任务的不确定性方面存在局限性。其核心原因在于,softmax分数实质上仅提供了检测分布的单点估计[15],且softmax输出往往对错误预测过于自信[16]。

为克服这一局限性,提出了证据深度学习(EDL)[17]。EDL基于Dempster-Shafer理论(DST)和主观逻辑(SL)的概念,为分类的不确定性提供了更为全面的技术。在K类分类任务中,给定K类分类的样本$x^{(i)}$,若假设类别概率服从先验Dirichlet分布时,学习证据$e^{(i)} \in R_+^K$涉及的交叉熵损失为以下形式:

$$Loss_{\mathrm{EDL}}^{(i)}(y^{(i)}, e^{(i)};\ \theta) = \sum_{k=1}^{K} y_k^{(i)}(\log S^{(i)} - \log(e_k^{(i)} + 1)) \tag{2}$$

本研究中$y^{(i)}$为K维独热编码标签,表示样本$x^{(i)}$的类别。证据$e^{(i)}$被定义为$e^{(i)} = g(f(x^{(i)};\ \theta))$。其中$e^{(i)}$由函数$g$和DNN的输出计算得出。$f$是由$\theta$参数化的判别器的输出,$g$为证据函数以保证证据$e^{(i)}$的非负性。Dirichlet分布为多变量概率分布,被定义为$Dir(p|\alpha)$,总强度被定义为$S = \sum_{k=1}^{K}\alpha_k$。根据Dempster-Shafer理论,$\alpha_k$通过等式$\alpha_k = e_k + 1$将$\alpha_k$与$e_k$联系起来。在推论中,第$k$类的预测概率为$\widehat{p_k} = \alpha_k/S$,预测不确定性$u$可以确定性地给出为$u = K/S$。

虽然基于EDL技术直接学习证据不确定性(无须依赖概率抽样)是可行的,但是在开放集入侵检测中分析未知CAN消息图像时,这种校准效果可能不理想。根据文献[18]和文献[19],校准良好的攻击检测模块在预测准确时应该对其预测充满信心,而在预测不准确时则应不确定。现有DNN模型的错误校准与负对数似然(NLL)的过拟合密切相关。此外,据文献[20],EDL函数(式(2))的目标是最小化NLL,这可能导致训练攻击检测模块过拟合,影响其在开放集入侵检测中的泛化性能。为解决这些问题,本文中基于准确度与不确定性间的动态关系校准EDL模型,通过最大化准确度与不确定性(accuracy versus uncertainty, AvU)效用函数,提出EDL模型的不确定性校准(EUC):

$$AvU = \frac{n_{\mathrm{AC}} + n_{\mathrm{IU}}}{n_{\mathrm{AC}} + n_{\mathrm{AU}} + n_{\mathrm{IC}} + n_{\mathrm{IU}}} \tag{3}$$

式中n_{AC}、n_{AU}、n_{IC}和n_{IU}分别代表4种检测情况下的样本数量:(1)准确且确定(AC);(2)准确但不确定(AU);(3)不准确但确定(IC);(4)不准确且不确定(IU)。校准良好的攻击检测模块应实现AvU的高效用,以确保检测不确定性与准确性相符。为校准检测的不确定性,EDL模型被鼓励学习准确且确定的预测或不准确且不确定的预测。因此,文中提出在训练阶段通过最小化AU和IC情况的期望,以规范EDL训练,鼓励AC和IU两种情况的发生。所以如果EDL模型对某数据点赋予高不确定性,则它更可能是不正确的,从而识别出未知入侵。通过考虑置信度p_i与不确定性u_i间的对数约束,最小化AU和IC情况的总和:

$$Loss_{\mathrm{EUC}} = -\lambda_t \sum_{i \in \{\hat{y}_i = y_i\}} p_i \log(1 - u_i) - (1 - \lambda_t) \cdot \sum_{i \in \{\hat{y}_i \neq y_i\}} (1 - p_i)\log(u_i) \tag{4}$$

式中:p_i代表输入CAN消息图像$x(i)$的最大类别概率;u_i代表相应的证据不确定性。第1项目标在攻击检测模块准确检测时($\widehat{y_i} = y_i, p_i \to 1$)实现低不确定性($u_i \to 0$),第2项目标则在攻击检测模块不准确检测时($\widehat{y_i} \neq y_i, p_i \to 0$)实现高不确定性($u_i \to 1$)。需要注意的是,退火因子$\lambda_t$(取值范围为$[\lambda_0, 1]$)定义为$\lambda_t = \lambda_0 \exp\{-(\ln \lambda_0/T)t\}$。$\lambda_t$是非负的,且初始值$\lambda_0$远小于1,$\lambda_t$随着训练周期$t$单调递增。当训练周期$t$增至总周期$T$时,$\lambda_t$将从初始值$\lambda_0$指数级递增至1.0。退火权重因子$\lambda_t$在攻击检测模块训练阶段起到动态平衡作用。在攻击检测模块训练阶段,准确和不准确检测主导时期不同。在训练的初期阶段,不准确检测通常占主导地位,此时第2项(IC损失)应加重惩罚;而在训练的后期阶段,准确检测通常占主导地位,因此第1项(AU损失)应加重惩罚。这种不确定度校准方法包括一个完全可微的正则化项,其优势在于训练过程中不依赖于可能出现分布偏移的验证集。这对于开放集入侵检测框架至关重要,因为开放集入侵检测框架通常无法接触到分布外(out-of-distribution, OOD)样本。因此,在校准大规模数据集时,本文中提出的框架提供了更高的灵活性。实验结果表明EUC方法在处理未知CAN消息图像类型时校准性能优势显著。

最终,在攻击检测模块训练阶段,应用式(1)中的EDL目标函数到CAN消息图像,该过程为每个数据类别收集"证据",有助于攻击检测模块中理解并区分不同类别CAN消息。在攻击检测模块测试阶段,假设数据类别的概率p遵循狄利克雷分布,即$p \sim Dir(p|\alpha)$,可以在一个$(K-1)$维的狄利克雷单纯形上同时表示数据的类别概率和不确定性(见图1中的三角形热图)。通过EDL的不确定性校准,使

得攻击检测模块能够更好地理解和量化其对CAN消息图像的不确定性,特别是对于未知的攻击类型,最终攻击检测模块将会输出预测的分类。

2　实验

2.1　实验设置

为全面评估本文提出的框架在性能优势和未知入侵检测方面的表现,并确保实验的公平性,将该框架与4种具有代表性的入侵检测方法(HyDL-IDS、LDAN、O-DAE和TSP)进行比较。评估指标采用精确度(Precision)、召回率(Recall)、F1分数(F1 Score)和准确性(Accuracy)4种经典指标。所有模块基于Python 3.7和pytorh1.11.0框架编写。训练和测试平台配备了CPU(Intel-6154)和GPU(TeslaV100s)。文中使用了两个数据集验证网络性能,其中数据集-1:汽车黑客数据集[18],是一个真实且广泛使用的公开数据集。该数据集从行驶中的车辆(车型为起亚Soul)OBD-II端口记录的实时CAN网络消息,涵盖正常CAN消息和4种攻击类型数据:DoS攻击、模糊攻击、伪造驾驶挡位攻击和伪造RPM仪表攻击,如表1所示。数据集-2:该数据集是通过奇瑞EXEED RX车辆在运行时通过OBD-II端口采集实时CAN消息来构建的,如图3所示。为收集CAN网络消息ID,一台CANalyst-II设备连接到车辆的OBD-II端口,从而能够捕获并记录CAN网络的数据流。采集的数据具体内容如表2所示。

表1　汽车黑客数据集

攻击类型	消息总数	正常消息	攻击消息
DoS攻击	3 665 771	3 078 250	587 521
Fuzzy攻击	3 838 860	3 347 013	491 847
GEAR攻击	4 443 142	3 845 890	597 252
RPM攻击	4 621 702	3 845 890	654 897
正常消息	988 987	988 987	0

图3　奇瑞EXEED RX车辆采集数据集

表2　奇瑞汽车数据集

攻击类型	消息总数	正常消息	攻击消息
DoS攻击	3 298 815	2 985 348	313 467
Fuzzy攻击	3 454 577	3 087 217	367 360
GEAR攻击	3 998 368	3 645 241	353 127
RPM攻击	4 159 054	3 904 785	254 269
正常消息	889 986	889 986	0

2.2　实验分析

2.2.1　已知攻击检测

实验将本模型与HyDL-IDS、LDAN、O-DAE等基准模型的性能进行比较,实验结果见表3。

表3　已知攻击检测

方法	攻击类型	Accuracy	Precision	Recall	F1 Score
HyDRL-IDS	Normal	0.997 5	0.983 5	0.980 5	0.982 0
	DoS	0.993 6	0.981 9	0.978 1	0.980 0
	Fuzzy	0.995 3	0.980 5	0.977 3	0.978 9
	GEAR	0.989 7	0.979 6	0.977 6	0.978 6
	RPM	0.989 5	0.981 3	0.981 3	0.980 7
LDAN	Normal	0.984 3	0.915 5	0.978 5	0.946 0
	DoS	0.980 6	0.909 9	0.975 6	0.941 6
	Fuzzy	0.980 2	0.912 4	0.971 3	0.940 9
	GEAR	0.981 4	0.920 1	0.976 4	0.947 4
	RPM	0.985 8	0.913 5	0.980 1	0.945 6
O-DAE	Normal	0.994 5	0.980 3	0.989 5	0.984 9
	DoS	0.993 3	0.974 2	0.984 3	0.979 2
	Fuzzy	0.991 2	0.987 23	0.983 9	0.978 1
	GEAR	0.987 9	0.965 3	0.978 9	0.974 2
	RPM	0.989 5	0.968 2	0.980 1	0.972 1
TSP	Normal	0.989 5	0.913 2	0.975 8	0.943 5
	DoS	0.980 2	0.910 0	0.972 8	0.940 4
	Fuzzy	0.981 1	0.912 5	0.973 3	0.941 9
	GEAR	0.980 0	0.902 8	0.967 8	0.943 2
	RPM	0.980 3	0.918 9	0.968 9	0.940 7
Ours	Normal	0.999 8	0.996 4	0.996 5	0.995 5
	DoS	0.998 7	0.995 4	0.991 5	0.995 8
	Fuzzy	0.999 2	0.994 9	0.993 2	0.994 0
	GEAR	0.997 9	0.994 5	0.993 5	0.994 2
	RPM	0.998 5	0.995 2	0.994 2	0.994 7

可以看出本文提出的方法与这些基准模型相比拥有最好的性能,在识别各种攻击类型时的精确度、准确性、召回率、F1分数均超过了0.99,证明了该方法在检测已知攻击类型时的有效性,但是这些基准模型都没有检测未知入侵的能力,文中将通过额外的实验证明提出的基于证据深度学习的方法以及证据不确定校准模块在检测未知攻击时的有效性。

2.2.2 未知攻击检测及消融实验

从训练数据集中排除一个类别的攻击消息不参与训练,本文中排除DoS消息,以便在测试时将DoS作为未知攻击样本以衡量模型未知攻击检测的能力。为验证证据深度学习方法对未知攻击检测能力的作用,文中与传统的基于softmax分类概率的方法进行对比,在应对训练集没有遇到过的攻击类型,传统的基于softmax分类概率进行未知攻击类型检测的做法是当某个样本最大的分类概率仍然小于某个阈值时,则将其分类为未知攻击类型,通过在我们的数据集上进行实验,得到的实验结果见表4。从表4中可以看出,本文方法拥有最佳的整体性能,尤其是在检测DoS这种未知攻击类型时,本文方法拥有99.87%的准确性,99.54%的精确度,99.15%的召回率,99.58%的F1分数,分别高于基于softmax的方法30.79%、21.09%、25.46%、19.76%,此外还分别将其他的攻击类型都排除在训练集之外以模拟未知攻击类型,实验结果(表5)表明,在各种未知攻击的检测上,提出的框架在各个指标上都显著高于传统基于softmax的分类方法,进一步证明文中提出的证据深度学习方法在检测训练阶段模型未学习到的样本类型时有显著的效果,同时还在本文模型中删除证据不确定性校准方法(EUC),并与本文提出的方法进行比较。实验结果表明,在5种消息类型上,有不确定性校准模块的方法能显著提高模型的性能,这得益于提出的方法最大化了准确性和不确定度的效用函数,可以鼓励模型学习准确且确定的预测或不准确且不确定的预测。

表4 未知攻击检测及EUC消融实验

类别	方法	Accuracy	Precision	Recall	F1 Score
Normal	基于softmax分类概率	0.933 4	0.967 2	0.976 5	0.983 0
	基于不确定度(无校准)	0.985 6	0.979 9	0.984 4	0.982 2
	基于不确定度(有校准)	0.999 8	0.996 4	0.996 5	0.995 5
DoS(未知)	基于softmax分类概率	0.690 8	0.784 5	0.736 9	0.798 2
	基于不确定度(无校准)	0.938 5	0.919 3	0.927 6	0.909 8
	基于不确定度(有校准)	0.998 7	0.995 4	0.991 5	0.995 8
Fuzzy	基于softmax分类概率	0.971 2	0.884 1	0.836 8	0.812 8
	基于不确定度(无校准)	0.985 5	0.923 4	0.965 4	0.945 6
	基于不确定度(有校准)	0.999 2	0.997 9	0.994 9	0.993 2
GEAR	基于softmax分类概率	0.993 0	0.997 1	0.994 3	0.994 5
	基于不确定度(无校准)	0.994 3	0.998 2	0.987 2	0.996 2
	基于不确定度(有校准)	0.997 9	0.994 5	0.993 5	0.994 2
RPM	基于softmax分类概率	0.996 0	0.953 5	0.972 5	0.983 6
	基于不确定度(无校准)	0.997 4	0.958 7	0.987 4	0.981 0
	基于不确定度(有校准)	0.998 5	0.995 2	0.994 2	0.994 7

表5 不同类型未知攻击检测

类别	方法	Accuracy	Precision	Recall	F1 Score
DoS(未知)	基于softmax分类概率	0.690 8	0.784 5	0.736 9	0.798 2
	基于不确定度(Ours)	0.998 7	0.995 4	0.991 5	0.995 8
Fuzzy(未知)	基于softmax分类概率	0.731 2	0.682 4	0.624 9	0.712 2
	基于不确定度(Ours)	0.998 1	0.997 9	0.994 9	0.993 2
GEAR(未知)	基于softmax分类概率	0.632 1	0.698 2	0.714 2	0.781 3
	基于不确定度(Ours)	0.996 1	0.987 1	0.992 2	0.998 3
RPM(未知)	基于softmax分类概率	0.652 0	0.621 8	0.712 2	0.703 5
	基于不确定度(Ours)	0.994 5	0.994 32	0.998 1	0.995 2

2.3 实验结果

使用黑客数据集和奇瑞数据集对提出的CAN网络入侵检测模型进行了实验,并将结果与几种常见的入侵检测基准模型进行对比。在检测已知攻击类型时,本文模型在各指标上都达到了99%,高于4种基准模型,本文提出的基于证据深度学习不确定度的入侵检测方法在面对未知入侵时效果显著,相比传统基于分类概率的方法性能指标平均提高了

24.5%,同时本文提出的证据不确定校准方法也可以有效校准模型,使得模型在预测时更加准确。

3 结论

提出了一种基于证据深度学习和生成对抗网络的车载CAN网络入侵检测框架,该框架专门设计用于有效检测已知和未知的攻击类别。在开源汽车黑客数据集和基于奇瑞EXEED RX车型自主构建的数据集上的检测结果显示了该模型在检测已知和未知攻击方面的有效性。相对于其他4个基准入侵检测框架,本文提出的方法始终表现出更优越的性能,相比于传统的基于softmax的未知入侵检测方法综合性能提高了24.5%。将证据深度学习集成到入侵检测中是本文采用的一种新颖方法,为解决车载CAN网络未知入侵检测提供了有效的解决方案。

参考文献

[1] JEONG H H, SHEN Y C, JEONG J P, et al. A comprehensive survey on vehicular networking for safe and efficient driving in smart transportation: a focus on systems, protocols, and applications[J]. Vehicular Communications, 2021, 31: 100349.

[2] ALIWA E, RANA O, PERERA C, et al. Cyberattacks and countermeasures for in-vehicle networks[J]. ACM Computing Surveys (CSUR), 2021, 54(1):1-37.

[3] 关宇昕,冀浩杰,崔哲,等.智能网联汽车车载CAN网络入侵检测方法综述[J].汽车工程,2023,45(6):922-935.
GUAN Y X, JI H J, CUI Z, et al. An overview of intrusion detection methods for in-vehicle CAN network of intelligent networked vehicles[J]. Automotive Engineering, 2023,45(6):922-935.

[4] AKSU D, AYDIN M A. MGA-IDS: optimal feature subset selection for anomaly detection framework on in-vehicle networks-CAN bus based on genetic algorithm and intrusion detection approach[J]. Computers & Security, 2022, 118: 102717.

[5] WEI P, WANG B, DAI X, et al. A novel intrusion detection model for the CAN bus packet of in-vehicle network based on attention mechanism and autoencoder[J]. Digital Communications and Networks, 2023, 9(1): 14-21.

[6] SWESSI D, IDOUDI H. Comparative study of ensemble learning techniques for fuzzy attack detection in in-vehicle networks[C]. International Conference on Advanced Information Networking and Applications. Cham: Springer International Publishing, 2022: 598-610.

[7] GROZA B, MURVAY P S. Efficient intrusion detection with bloom filtering in controller area networks[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(4): 1037-1051.

[8] HOANG T N, KIM D. Supervised contrastive ResNet and transfer learning for the in-vehicle intrusion detection system[J]. Expert Systems with Applications, 2024, 238: 122181.

[9] SUN Y, OCHIAI H, ESAKI H. Intrusion detection with segmented federated learning for large-scale multiple lans[C]. 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-8.

[10] MCHERGUI A, MOULAHI T, ZEADALLY S. Survey on artificial intelligence (AI) techniques for vehicular ad-hoc networks (VANETs)[J]. Vehicular Communications, 2022, 34: 100403.

[11] KIM K, KIM J S, JEONG S, et al. Cybersecurity for autonomous vehicles: review of attacks and defense[J]. Computers & Security, 2021, 103: 102150.

[12] WU W, LI R, XIE G, et al. A survey of intrusion detection for in-vehicle networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(3): 919-933.

[13] LIANG J, CHEN J, ZHU Y, et al. A novel Intrusion Detection System for Vehicular Ad Hoc Networks (VANETs) based on differences of traffic flow and position[J]. Applied Soft Computing, 2019, 75: 712-727.

[14] MUSA U S, CHAKRABORTY S, ABDULLAHI M M, et al. A review on intrusion detection system using machine learning techniques[C]. 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, 2021: 541-549.

[15] GAL Y. Uncertainty in deep learning[D]. University of Cambridge, 2016.

[16] GUO C, PLEISS G, SUN Y, et al. On calibration of modern neural networks[C]. International Conference on Machine Learning. PMLR, 2017: 1321-1330.

[17] SENSOY M, KAPLAN L, KANDEMIR M. Evidential deep learning to quantify classification uncertainty[J]. Advances in Neural Information Processing Systems, 2018, 31.

[18] MUKHOTI J, GAL Y. Evaluating bayesian deep learning methods for semantic segmentation [J]. arXiv preprint arXiv: 1811.12709, 2018.

[19] KRISHNAN R, TICKOO O. Improving model calibration with accuracy versus uncertainty optimization[J]. Advances in Neural Information Processing Systems, 2020, 33: 18237-18248.

[20] MUKHOTI J, KULHARIA V, SANYAL A, et al. Calibrating deep neural networks using focal loss[J]. Advances in Neural Information Processing Systems, 2020, 33: 15288-15299.