AI大模型赋能网络流量分类概述

Overview of AI Big Model Empowering Network Traffic Classification

陈雪娇[1]，付梦艺[2]，王　攀[2]（1. 南京信息职业技术学院，江苏 南京 210023；2. 南京邮电大学，江苏 南京 210003）
Chen Xuejiao[1], Fu Mengyi[2], Wang Pan[2]（1. Nanjing Vocational College of Information Technology, Nanjing 210023, China; 2. Nanjing University of Posts and Telecommunications, Nanjing 210003, China）

摘　要：
提出一个通用的AI驱动的网络流量分类框架，阐述了所涉及的工作流程、分类目标、设计原则以及典型场景等，并提出了一个基于BERT的网络流量分类模型，通过将输入的分组净荷进行向量化嵌入，然后送入BERT进行预训练，用于实现流量数据的上下文理解并捕获双向特征，然后对接一个全连接网络对分类下游任务进行微调，从而实现流量分类。通过与AE、VAE、ByteSGAN 3个经典的流量分类深度学习模型在CICIDS2017公开数据集上进行对比，发现BERT的精度明显高于其他方法。

关键词：
流量分类；流量识别；入侵检测；BERT；大模型
doi：10.12045/j.issn.1007-3043.2024.09.003
文章编号：1007-3043（2024）09-0013-07
中图分类号：TN915.5
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
It introduces a comprehensive AI-driven framework for network traffic classification, delineating the workflow, classification objectives, design principles, and typical application scenarios. Additionally, it proposes a BERT-based model for network traffic classification by leveraging packet payload vectorization and embedding it into BERT for pre-training to achieve contextual comprehension of traffic data and capture bidirectional features. Subsequently, fine-tuning is conducted using a fully connected network to accomplish traffic classification tasks. Comparative analysis with three classical traffic classification deep learning models（AE, VAE, and ByteSGAN）on the CICIDS2017 public dataset demonstrates that BERT achieves significantly higher accuracy than other methods.

Keywords：
Network traffic classification; Traffic identification; Intrusion detection; BERT; Big model

引用格式：陈雪娇，付梦艺，王攀. AI大模型赋能网络流量分类概述[J]. 邮电设计技术，2024（9）：13-19.

0 引言

作为网络管理和安全的重要手段，网络流量分类（Network Traffic Classification，TC）自上世纪90年代末开始就得到学术界和工业界的高度关注，在QoS/QoE管理、网络资源优化、拥塞控制、入侵检测等方面都取得了很好的应用。随着新一代网络技术（B5G/6G、物联网、天地一体化网络等）的快速发展，网络技术正朝着“自愈、自管理、自优化和自保护”的高度自治化方向发展，网络流量分类技术作为精细化网络业务和安全管理的决策手段之一，扮演着关键角色。然而随着海量异构终端的泛在接入，网络呈现出高度的“动态性”“异质性”和“复杂性”，这给网络流量分类技术带来了一系列新的挑战。

TC技术的发展大致经历了3个阶段。第1阶段基于端口/DPI实现TC，然而随着越来越多的应用采用隧道、加密、随机端口等技术，加之涉及用户隐私泄露等

基金项目：国家自然科学基金（61972211）
收稿日期：2024-08-16

安全问题，这类技术很快失效[1]。第2阶段主要采用机器学习（Machine Learning，ML）、概率统计等方法，包括SVM、RF、DT、KNN等[2]。然而，这类方法需要提取高质量的流量特征作为ML的训练基础，而这些特征的提取和选择高度依赖于网络专家的经验，且费时费力，无法满足网络和业务的快速演进和发展，从而造成“慢半拍”现象。此外，网络流量数据的“海量性”使得基于ML的TC方法在训练和分类方面不堪重负，难以满足工业界的实际应用需求。随着云计算、大数据，尤其是深度学习（Deep Learning，DL）和高性能计算技术的高速发展，海量流量数据的特征学习成为可能，给TC领域带来了新的提升空间。2015年，王占一等人[3]首次提出采用卷积神经网络（CNN）、堆栈式自动编码机（Stack Auto-Encoder）等DL模型实现流量分类，使TC技术发展进入第3阶段。DL有3个优点：自动提取特征、可揭示更深层次的数据规律和大量成熟应用于计算机视觉/图像/文本/语音的模型可复用，这些优点恰好是基于ML的TC方法所欠缺的，自此，基于DL的TC分类技术（下文简称DL-TC，后文中的AI-TC指ML/DL-TC）迎来了新一波的热潮，一系列的DL-TC分类方法被提出，包括基于CNN/AE/MLP/LSTM/GAN等方法，并取得了比ML-TC算法更好的分类性能。随着大语言模型（Large Language Model，LLM）的出现，其优异的内容生成能力给通信网络领域的研究者带来了全新的思路，本文将Transformer、BERT以及LLM赋能于网络流量分类定义为TC的第4次浪潮。

尽管DL-TC的研究工作取得了一系列成果，但在工业界（比如运营商、工/企业网等）始终未被实际应用，笔者认为现有的AI-TC技术仍存在诸多局限性。

a）数据集问题。数据集是AI模型的基础，而现有AI-TC模型训练普遍采用公开数据集，这些公开数据集往往“量少、过时、质量无法考证”。

b）资源受限条件下的模型轻量化问题。AI-TC如何在网络边缘设备（如物联网网关、家用路由器、5G CPE等）乃至一些弱计算能力的终端上实现推理/分类功能。

c）成本问题。在训练和推理阶段，计算资源（处理器/内存/Flash）、时间、人力等成本消耗与分类性能之间如何求得平衡。

d）可信问题。如何解决AI-TC模型的“黑盒子”问题，让分类模型的使用者（比如运营商）信任模型。

e）演进问题。如何解决因业务/应用/攻击的“快速演进”而造成的分类模型“慢半拍”以及“道高一尺、魔高一丈”问题，比如新应用、Zero-day攻击、“流变种”等。

f）数据/模型隐私问题。如何防范数据集的敏感数据泄露以及攻击者对分类模型的反推解构乃至对分类模型实施攻击等问题。

本文针对以上AI-TC所面临的挑战，提出一个通用的端到端AI-TC的工作流程；并给出AI-TC的需求和设计原则的定义、应用场景；然后围绕AI-TC的工作流程，细化并总结当前面临的各项挑战及研究进展；最后提出AI大模型赋能网络流量分类的设想以及存在的困难。

1 一个通用的端到端AI-TC分类工作流程

为了应对“网络环境动态变化、业务应用的快速进化、隐私保护持续升级”的三大挑战，AI-TC分类系统的工作流程必须是一个持续学习的迭代优化过程。从一个机器学习的全生命周期角度来看，一个通用的端到端AI-TC分类工作流程分为2个阶段：部署前和部署后（见图1）。部署前阶段主要分为AI-TC分类需求/设计原则的定义和模型开发与评估，部署后主要是模型的推理阶段。

a）需求定义主要用于定义AI-TC的详细分类需求，包括分类颗粒度、应用场景、数据来源、实时/非实时等；设计原则主要用于定义该分类系统的开发原则，包括可靠性、鲁棒性、安全性、自适应性等。

b）模型开发和评估主要包括数据工程、特征工程、模型训练、模型评估/解释和模型部署，该过程依据上一阶段定义的需求和设计原则，训练和输出高性能的AI-TC模型，值得注意的是，整个过程是一个不断迭代动态优化的过程，而不是静态的，其迭代优化的触发通过2个途径：第一，通过模型解释环节对当前输出的分类模型进行透明性、公平性和可信度分析，给数据工程、特征工程和模型训练提供优化依据；第二，通过模型推理环节的性能监测，不断给出模型当前的分类实时性能，及时向模型部署环节报告性能劣化事件，从而驱动模型的新一轮迭代优化。

c）在模型推理阶段，AI-TC分类模型实际服务于不同的应用场景，根据需求的不同，可以部署于网络侧、边缘侧和终端侧，服务于多种不同的分类任务。面对网络环境的动态性和业务应用的演进变化，在推理环节不但要做好分类任务的执行，还需要通过性能

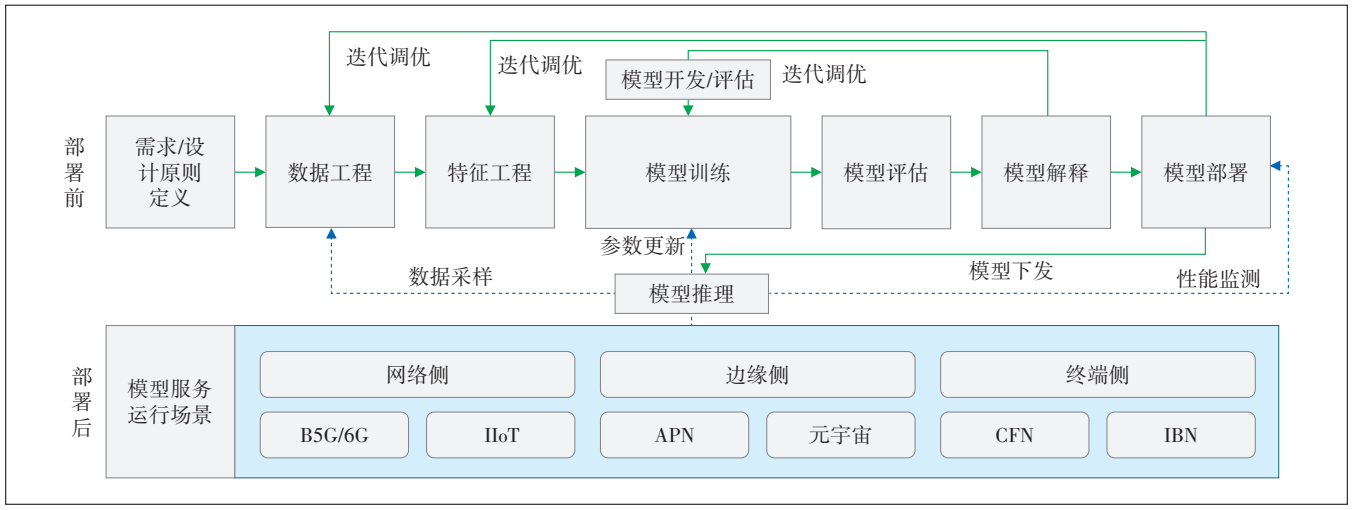


图1　通用端到端AI-TC工作流程

监测，及时发现和上报分类性能劣化事件，从而驱动模型的新一轮迭代优化。在此期间，分类系统需根据设定好的采样策略，不断向数据工程模块输入具有代表性的新数据，以便形成新的训练数据集。同时，分布式TC场景下，为保护用户的上网行为隐私，AI-TC分类器需在网络边缘/终端上进行本地训练，并向集中训练节点更新参数。

2 AI-TC的分类目标、设计原则和典型场景

2.1 分类目标

定义和明确一个AI-TC的分类目标，需要从网络环境、应用场景、分类粒度、实时性和轻量化5个方面来考虑。

a）网络环境。即流量分类的目标网络对象，不同的网络环境对流量分类的要求不同。传统的网络环境包括宽带互联网、蜂窝无线网络（3G/4G）、Wi-Fi、MANET等；新一代网络环境包括B5G/6G、确定性网络、IIoT、天地空一体化网络等。

b）应用场景。即流量分类的应用目标，总体而言，AI-TC是为业务/应用管理和网络安全2种场景服务的。细分而言，业务/应用管理包括QoS/QoE管理、网络资源管理和优化、拥塞控制、网络资产发现等；网络安全包括入侵检测、恶意流量识别、网络取证等。随着加密、VPN隧道、匿名化技术（比如Tor）、流量突变（Traffic Mutation）、伪装等流量混淆技术（Traffic Obfuscation）的不断演进，AI-TC在网络安全场景方面面临严峻的挑战。

c）分类粒度。主要分为粗粒度（Coarse-grained）、细粒度（Fine-grained）和二分类3种。协议/业务识别属于粗粒度，应用/设备/OS识别属于细粒度，正常/异常流量分类属于二分类。

d）实时性。主要分为在线（Online）和离线（Offline）。前者往往应用于管理和控制类的应用场景，而后者则用于分析、检测和取证等场景。

e）轻量与否。根据AI-TC分类器所需的计算资源是否受限，从而明确分类器是否需要按照轻量级标准设计。很多应用场景希望在网络边缘设备甚至弱计算终端上执行分类任务，由于计算资源有限，AI-TC分类模型必须是轻量级的。

2.2 设计原则

针对网络环境的“动态性”“异质性”“资源受限”和“业务/应用/攻击快速演进”四大特点，本文提出一个轻量级的AI-TC分类器所需遵循的七大设计原则，以满足多样化的需求（Requirement，RQ），其中包括可靠性（Reliability）、OR高效性（Efficiency）、健壮性（Robustness）、扩展性（Scalability）、安全性（Security）、可解释性（Interpretability）、适应性（Adaptability）、实时性（Realtime）以及成本（Cost）。

2.3 应用场景

图2所示为一个网络边缘侧的轻量级AI-TC应用场景。如前所述，海量异构终端（传感器/PLC/CNC/VR/AR/可穿戴设备等）以泛在接入的方法连接到多样化的网络环境中。这些边缘设备包括5G CPE、IIoT工业网关、智慧家庭网关/家用路由器等，承担着异构终端接入、协议转换、数据转发、边缘智能等功能，边缘设备所要处理的流量中混杂着传感数据、控制数据、

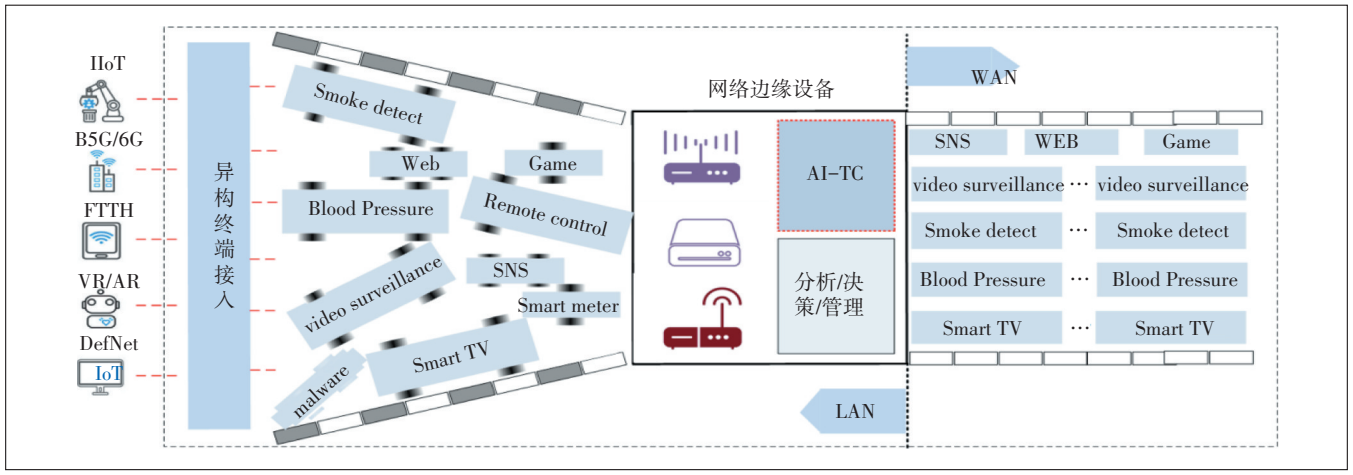


图2　网络侧AI-TC应用场景

音/视频、VR/AR乃至恶意攻击的流量。为了确保关键业务的QoS/QoE管理、资源优化、拥塞控制和安全防御，流量分类成为网络管理者分析、决策和实施管理的前提。随着网络管理向智能化、自主化的逐步发展，AI成为了关键的赋能者，AI-TC成为了网络设备边缘智能的一个关键能力。

现有大多AI-TC研究工作聚焦于传统的宽带互联网、蜂窝无线网络（3G/4G）、Wi-Fi等，近些年涌现出的新一代网络技术对AI-TC也提出了迫切的需求。

a）AI驱动自治网络业务管理。B5G/6G提出自治化（Autonomous Network）网络[4]，包括自管理、自愈、自优化和自保护等机制，其目的就是实现网络和业务管理的高度自治化。很多标准化组织均提出了网络自主化管理的框架/标准，其中，ETSI提出的ZSM（Zero-touch network and Service Management）框架旨在利用AI驱动网络管理决策，进而优化CAPEX和OPEX，得到了越来越多工业界和学术界的关注。其中，对泛在接入的海量网络流量的分类识别是一项必不可少的基础工作，而AI技术成为了关键的赋能者。

b）端到端确定性通信。确定性网络（Deterministic Network，DetNet）作为支撑工业互联网（Industrial Internet of Things，IIoT）的一个核心网络技术，可以为IIoT业务/应用提供确定性的时延、抖动和丢包率等。时间敏感网络（Time-Sensitive Network，TSN）是一种主流的确定性网络技术，但目前的TSN大多局限于封闭工厂内网/域。而多数工业互联网应用是广域场景，需通过多个TSN网络，且跨越非TSN网络实现端到端通信。工厂内网连接了诸多工厂生产要素，如PLC、CNC、传感器等，如何在工业内网边缘侧精准识别和分类工业内网应用流量，并联动SDN控制器对多个TSN网络或非TSN网络进行全局业务编排和统一资源管理，是实现准确高效的跨网跨域端到端确定性通信，进而实现工业互联网全连接服务质量保障的关键。

c）B5G/6G精细化网络切片划分。当前研究主要针对大类业务的粗粒度网络切片，而未来B5G/6G业务会存在更细粒度的网络切片划分需求，用以支持面向计算的通信、上下文敏捷的eMBB通信和事件定义的uRLLC增强型业务，形成沉浸式云扩展现实、全息通信、感官互联、智慧交互、通信感知、普惠智能、数字孪生、全域覆盖等业务应用。

d）应用感知的IPv6网络（Application-awareness IPv6 Network，APN6）[5]。APN6通过数据分组携带应用的标识和需求信息，即传递应用信息，并保证应用信息的安全可靠，使网络感知应用并根据应用的需求为其提供优质的差异化服务。而网络在传送数据分组时，根据数据分组中的应用信息匹配网络对应策略，并选择相应的SRv6路径传输数据分组（如低时延路径），满足SLA需求，提高服务质量。

3 AI大模型赋能网络流量分类

无论是经典模型，还是大模型，应用于网络流量分类领域均需要考虑网络数据的表达学习和强泛化分类算法2个重要问题。

3.1 网络数据的表达学习（Representation Learning）

通信网络“结构复杂、业务多样”，导致其信息多为富媒体，且来自于网管、CRM、业务、安全等支撑类的数据更多，这些数据包括文本、图像、语音、视频、时序、数据包、路由记录、拓扑结构等多种混杂形式，具

有结构化/非结构化的信息结构。至今，网络感知数据没有合适的统一信息表达方式，更无适合的神经网络范式与之相匹配。将DL应用于网络乃至于TC领域的技术大多采用将网络数据转换为传统模态[6]，比如图像、视觉或文本，再套用与之相匹配的深度学习算法建模，这种硬性的匹配导致模型的泛化能力始终不佳。

近年来，有部分学者提出采用图对网络数据进行表达。网络数据，包括流量、拓扑、QoS等，属于典型的非欧几里得数据，且富含大量的时序关系，采用传统的多维向量和矩阵已很难表达其复杂的时序及空间关系[7]。而GNN具有如下3个重要的特征：可以处理非欧几里得空间的数据；模型输入的形状灵活，不像CNN/LSTM需要固定形状；图中的节点可以共享全局信息。尤其是第2条，一旦输入形状灵活，那么模型的特征就可以不再受CNN/LSTM模型输入的限制，也即为了固定形状，必须对输入特征进行固定长度的截断（Truncation）或补零（Zero-padding），从而造成信息丢失或冗余；然而，采用图表达的网络数据通过GNN的建模后，泛化能力始终不理想。随着基础模型（Foundation Model，FM）的出现，尤其是基于FM的大语言模型（Large Language Model，LLM），在泛化能力和生成能力上异军突起，研究人员开始探索基于FM的网络流量分类的研究。

3.2 基于BERT/Transformer的网络流量分类模型

大型语言模型（LLM）最近在自然语言处理、计算机视觉等领域得到了快速发展。在网络流量分类中，Hendrycks 等人的研究表明，尽管预训练可能不会提高传统分类指标的性能，但它提高了模型的鲁棒性和不确定性估计。文献[8]通过对标签损坏、类别不平衡等进行大量实验，证明了预训练带来的巨大收益以及与任务特定方法的互补效果。NetGPT首次尝试为流量理解和生成任务提供生成式预训练模型。通过多模式网络流量建模，统一了文本输入、报头字段清理、数据包分割以及标签和提示合并，以优化预训练模型对各种任务的适应性[9]。Horowicz 等人提出使用无监督对比学习（Contrastive Learning，CL）来增强流量图像样本，并缓解小样本流量分类问题[10]。Lin等人提出了一种称为ET-BERT的新流量特征表示模型，该模型从大规模原始流量数据包中提取上下文数据包级表示，以提高下游分类任务的准确性[11]。Ferrag 等人[12]提出了一种使用大型语言模型的新型基于网络的网络威胁检测方法，并引入了一种称为固定长度语言编码（FLLE）的隐私保护编码方法。该论文从头开始实现并训练用于多类别分类的BERT架构，采用FalconLLM作为事件响应和恢复系统。然而，考虑到原始流量输入和网络任务输出的独特特性，为网络流量构建BERT预训练模型仍然面临以下几个挑战：由于流量数据中的报头和有效负载具有异质性，如何有效地整合语义信息具有挑战性；如何设计有效的预训练任务以实现流量数据的上下文理解并捕获双向特征至关重要；在网络安全等关键领域，如何解释BERT等大型语言模型的决策过程和输出结果对于信任和合规性是有必要的。

图3所示为本文提出的一个基于BERT（Bidirectional Encoder Representations from Transformers）的预训练模型用于网络流量分类的流程。

a）数据预处理。将输入的数据包载荷（facb34...aebb）进行字节切分（如fa、cb等），切分后的单词被作为输入载荷的最小单元，经过字节合并后的载荷将被作为BERT模型的输入。

b）输入嵌入。每个数据包载荷会经过词嵌入、段落嵌入和位置嵌入3个步骤来生成向量表示。词嵌入是将词汇映射到高维空间中的数值表示；段落嵌入考虑了单词在段落中的上下文信息；位置嵌入则反映了单词在序列中的相对位置关系。

c）BERT微调。这些嵌入向量会被送入预训练的BERT模型进行进一步处理。BERT模型由12个Transformer编码器组成，每个Transformer编码器包含多头注意力机制（Multi-Head Attention）[13]和前馈神经网络（Feed Forward Network），它们共同作用于输入数据以提取更高级别的特征。BERT模型的预训练任务通常由掩码语言任务和下一句预测任务构成，通过大量未标注的流量包载荷数据来优化这2个任务的损失函数。

d）网络流量分类。提取[CLS]特殊标记对应的输出向量作为最后一层全连接层的输入用于下游任务。根据下游任务需要，可以使用[CLS]对应的特征向量来进行包级流量分类或者流级流量分类。具体来说，对于包级流量分类，可能只需要根据单个数据包的内容做出决策；而对于流级流量分类，则可能需要考虑连续多个数据包之间的关联性才能准确判断其类别，例如拼接数据包或加入一条流的数据包长度信息等。

图4所示为4种不同的方法（AE、VAE、ByteS-

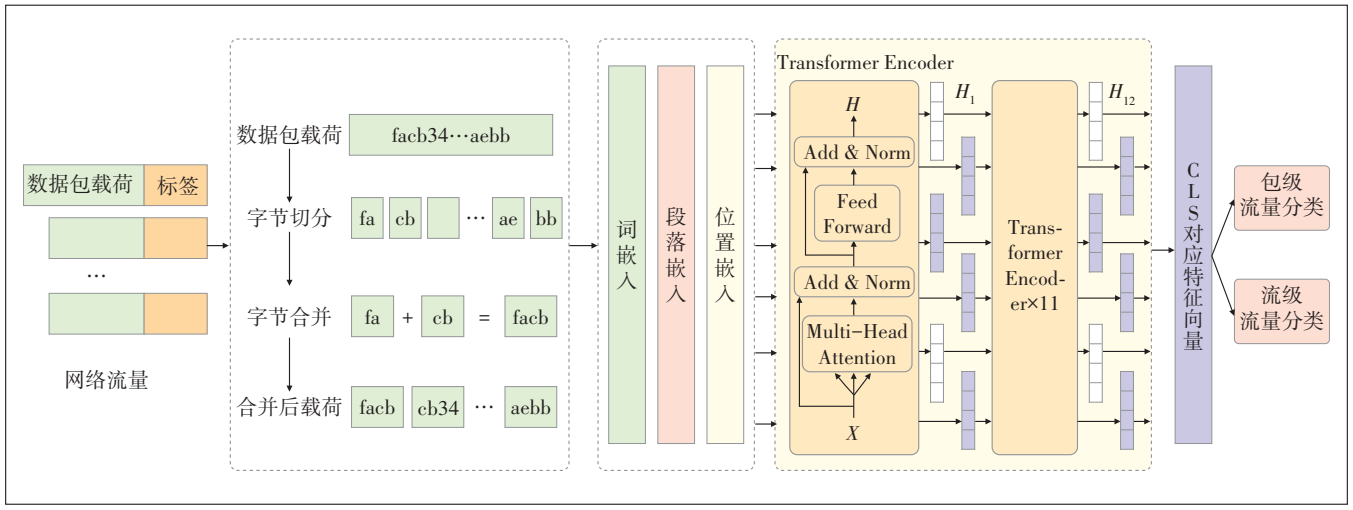


图3　基于BERT的网络流量分类模型

GAN[14]和BERT)在CICIDS2017[15]数据集上的分类性能表现。横坐标表示不同的攻击类型，纵坐标表示指标。从图4可以看出，对于大多数攻击类型，BERT的分类性能都在其他3种方法之上，这表明BERT在识别多种类型的网络攻击方面通常具有较高的准确性。尤其是在识别“A8:Slowloris”和“A9:Slowhttptest”这2种攻击时，BERT的精度明显高于其他方法，说明它在这2个特定场景下的表现非常出色。

4 结束语

本文通过对AI赋能TC这一研究方向的历史沿革、典型场景的描述，提出了一个端到端的通用AI-TC工作流程，并进一步深入研究AI大模型如何赋能网络流量分类，提出了一个基于BERT的网络流量分类模型，通过将输入的分组净荷进行向量化嵌入，然后送入BERT进行预训练，用于实现流量数据的上下文理解并捕获双向特征，然后对接一个全连接网络对分类下游任务进行微调，从而实现流量分类。与AE、VAE、ByteSGAN这3个经典的IDS深度学习模型在CICIDS2017公开数据集上的对比结果表明，BERT的精度明显高于其他方法。然而，如何解释BERT等大型语言模型的网络流量分类的决策过程和输出结果对于信任和合规性是非常重要的，这将是我们未来的研究方向。

参考文献：

[1] 申进. 基于DPI和DFI的网络流量分类方法研究与应用[D]. 绵阳：西南科技大学，2020.

[2] SHEN M, YE K, LIU X T, et al. Machine learning-powered encrypted network traffic analysis: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 791-824.

[3] WANG Z Y. The applications of deep learning on traffic identification [EB/OL]. [2024-01-21]. https://www. blackhat. com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification-wp.pdf.

[4] PAPIDAS A G, POLYZOS G C. Self-organizing networks for 5G and beyond: a view from the top[J]. Future Internet, 2022, 14(3): 95.

[5] PENG S P, MAO J W, HU R Z, et al. Demo abstract: APN6: application-aware IPv6 networking [C]//IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE, 2020: 1330-1331.

[6] 周子云，黄洪. 改进EfficientNet图像分类的恶意流量检测模型[J]. 四川轻化工大学学报(自然科学版)，2023，36(6)：49-56.

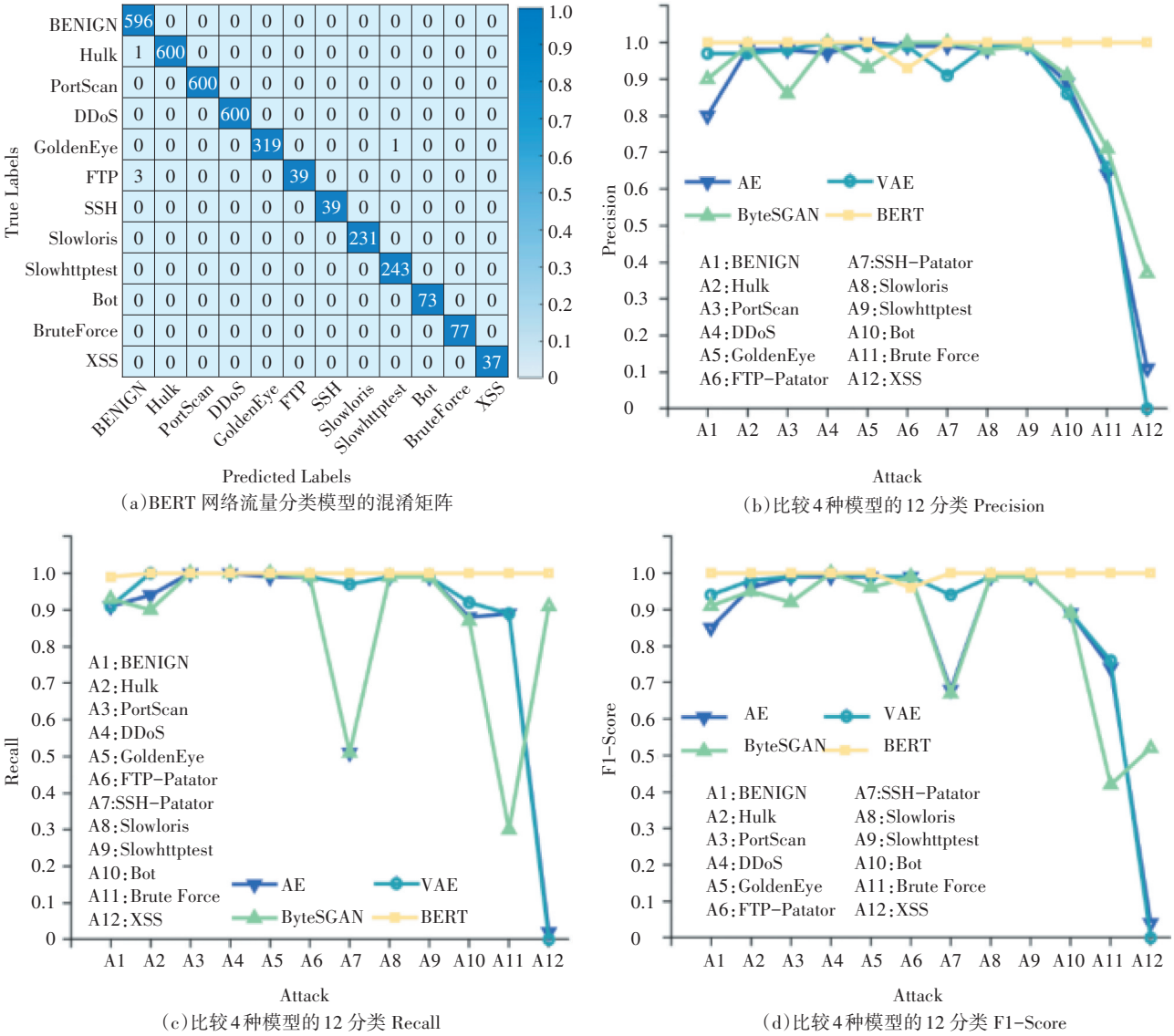
[7] HUOH T L, LUO Y, LI P L, et al. Flow-based encrypted network traffic classification with graph neural networks[J]. IEEE Transactions on Network and Service Management, 2023, 20(2): 1224-1237.

[8] HENDRYCKS D, LEE K, MAZEIKA M. Using pre-training can improve model robustness and uncertainty[C]//Proceedings of the 36th International Conference on Machine Learning. New York: PMLR, 2019: 2712-2721.

[9] MENG X Y, LIN C G, WANG Y Q, et al. Netgpt: generative pre-trained transformer for network traffic [EB/OL]. [2024-01-22]. https://arxiv.org/pdf/2304.09513.

[10] HOROWICZ E, SHAPIRA T, SHAVITT Y. A few shots traffic classification with mini-flowPic augmentations [C]//Proceedings of the 22nd ACM Internet Measurement Conference. New York: Association for Computing Machinery, 2022: 647-654.

[11] LIN X J, XIONG G, GOU G P, et al. ET-BERT: a contextualized datagram representation with pre-training transformers for encrypted traffic classification [C]//Proceedings of the ACM Web Conference 2022. New York: Association for Computing Machinery, 2022: 633-



(a)BERT 网络流量分类模型的混淆矩阵

(b)比较4种模型的12分类 Precision

(c)比较4种模型的12分类 Recall

(d)比较4种模型的12分类 F1-Score

图4　基于BERT的网络流量分类模型性能评估

642.

[12] FERRAG M A, NDHLOVU M, TIHANYI N, et al. Revolutionizing cyber threat detection with large language models: a privacy-preserving BERT-based lightweight model for IoT/IIoT devices[J]. IEEE Access, 2024(12): 23733-23750.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017: 6000-6010.

[14] WANG P, WANG Z X, YE F, et al. ByteSGAN: a semi-supervised generative adversarial network for encrypted traffic classification in SDN edge gateway[J]. Computer Networks, 2021(200): 108535.

[15] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy - ICISSP. Setúbal: SciTePress, 2018: 108-116.

作者简介：

陈雪娇，副教授，硕士，主要研究方向为信息网络、深度学习、信息安全、B5G/6G通信以及深度学习在网络安全中的应用；付梦艺，博士，主要研究方向为新一代信息网络、深度学习、信息安全、B5G/6G通信；王攀，博士生导师，研究员，主要研究方向为新一代信息网络、深度学习、信息安全、B5G/6G通信。