计算机科学
COMPUTER SCIENCE

一种噪声容忍的网络流量分类方法

马继烨, 朱国胜, 卫操, 曾堉萱

引用本文

马继烨, 朱国胜, 卫操, 曾堉萱. 一种噪声容忍的网络流量分类方法[J]. 计算机科学, 2023, 50(11A): 220800120-7.

MA Jiye, ZHU Guosheng, WEI Cao, ZENG Yuxuan. Noise Tolerant Algorithm for Network Traffic Classification Method [J]. Computer Science, 2023, 50(11A): 220800120-7.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220800120

一种噪声容忍的网络流量分类方法

马继烨　朱国胜　卫　操　曾堉萱
湖北大学计算机与信息工程学院　武汉 430062
(majiyemjy@126.com)

摘　要　针对传统基于机器学习的网络流量分类方法中样本标签的正确性会直接影响结果精度的问题,提出一种噪声容忍的网络流量分类方法。该方法基于深度残差网络的方法,首先,对网络流量数据进行归一化以及数据增强处理后映射成灰度图片,并对其样本标签进行不同程度的加噪;然后,基于 Res2Net 深度残差神经网络设计适用于网络流量噪声干扰下的维度模块,构造可以适用于流量标签噪声容忍的深度神经网络模型。基于公开数据集的实验结果表明,与传统的噪声容忍分类算法相比,基于改进的深度残差神经网络在不同噪声率下均提升了分类精度,并且在高噪声率下提升更为显著。

关键词:噪声容忍;深度学习;残差学习;流量分类;标签噪声;归一化
中图法分类号　TP393

Noise Tolerant Algorithm for Network Traffic Classification Method

MA Jiye,ZHU Guosheng,WEI Cao and ZENG Yuxuan
School of Computer and Information Engineering,Hubei University,Wuhan 430062,China

Abstract　Aiming at the problem that the correctness of the sample labels in the traditional machine learning-based network traffic classification method will directly affect the accuracy of the results,a noise-tolerant network traffic classification method is proposed,which is based on the deep residual network method. After normalization and data enhancement,the data is mapped into a grayscale image,and the sample labels are added to different degrees of noise. Then,based on the Res2Net deep residual neural network,a dimensional module suitable for the interference of network traffic noise is designed,and a deep neural network model suitable for traffic label noise tolerance is constructed. Experimental results on public datasets show that compared with the traditional noise-tolerant classification algorithm,the improved deep residual neural network improves the classification accuracy under different noise rates,and the improvement is more significant at high noise rates.

Keywords　Noise tolerant,Deep learning,Residual learning,Network traffic classification,Label noise,Normalized

1　引言

准确的网络流量分类作为众多网络活动的基础,在网络管理和网络安全等领域都非常重要。随着互联网的发展以及新型应用的不断兴起,通常的网络分类方法由于加密技术的推广、应用程序的数量和类型的不断增长,传统基于端口、报文内容的流量分类方法已经不再对所有类型的网络流量有效[1]。

基于深度学习的流量分类方法一直是研究者关注的热点。文献[2]提出一种基于卷积神经网络的流量分类算法,其基于改进的 CNN 流量分类方法,不仅提高了流量分类的精度,而且减少了分类所用的时间。文献[3]提出了一种基于 GADCN 的流量识别模型,用于识别良性流量和恶意软件流量,解决了灰度图像质量低和数据集类别不平衡的问题,提高了恶意软件流量的识别精度。文献[4]提出了一种基于特征融合的轻量级网络模型 Inception-CNN,用于端到端加密流量的分类,在显著提高分类结果准确性的同时,降低了网络计算复杂度。文献[2-4]均采用转换成图像进行图像识别的方式进行分类。

虽然深度神经网络在大规模数据集上的图像分类中表现出了卓越的性能,但是这些模型通常对数据集中标签类别的准确性有着极高的要求。网络上收集的数据往往包含不准确的标签,这些噪声被称为噪声标签[5]。传统的深度神经网络可以学习和训练任何数据集,但在有噪声的数据集下深度神经网络很容易会过拟合,甚至可能会记住噪声[6]。对于真实的数据集,其中的多数样本会在一定的训练后被正确分类,这些样本被称为简单类。而少数样本的分类准确率并不会随着训练的增加而提升,这些样本被称为困难类。文献[7]进一步对此进行验证,认为在真实的数据集中,传统的深度神经网络优先考虑简单分类的学习,然后才会拟合噪声。在数据从拟合真实数据转移到拟合噪声的过程中,临界采样率会大大增加,意味着需要更多数据去解释噪声,因此真实数据集更容易拟合到噪声标签中。对大规模的数据集进行标签标记意味着大量的人力以及相当高的专业知识,这大大提高了数据集的成本。人工标记标签通常会产生噪声标签,且噪声标签更偏向于集中在困难类样本中,该真实标签类别的样本与噪声标签类别样本的相似性导致了困难类的分类难度进一步提升。与只使用少量干净标签的数据集相比,使用大量的但是带有噪声标签的数据集进行训练的效果更好[8]。在有噪声标签的情况下训练精确的深度神经网络,是深度学习

通信作者:朱国胜(zhuguosheng@hubu.edu.cn)

中一项非常重要的实际任务。

2 标签噪声数据集问题及描述

针对噪声标签的学习分为基于无噪声模型方法和基于噪声模型方法。基于噪声模型方法一般有清洗标签和数据集删减等。文献[9]提出一种噪声标签校准方法,将预测噪声标签与预测标签的置信度的似然比与预定阈值进行比较,当置信度小于阈值时对噪声标签进行清洗,以逐渐提升模型的性能。该方法要求其在未经清洗前的错误率低于噪声率,否则在随机翻转标签后数据集的噪声率会明显增加。文献[10]通过调整学习率,使网络在欠拟合和过拟合之间变化,存在噪声标签的样本的损失值更大,因此在训练中删除了部分较大的噪声标签样本,以抑制噪声标签的负面影响。该方法可以有效删除简单类噪声,然而在困难类噪声中效果并不明显。文献[11]首先在噪声数据上训练网络,并使用该模型提取特征向量;然后使用基于提取特征的 K-means 算法对数据进行聚类,并去除异常值。基于噪声模型的方法很大程度上依赖于对噪声结构的准确分析,通常对数据噪声标签的分布、特点做出假设,这损害了不同噪声标签的设置的适用性。

基于无噪声模型方法一般有改编损失函数和元学习等。文献[12]通过非完全对称学习将损失函数进行改编,使得模型在噪声数据集下具有更好的鲁棒性,并且对现有模型架构的改动较小,可以快速应用。文献[13]在传统梯度更新之前执行元学习更新,在元训练中生成多个小批次的带有合成噪声标签的样本并用其更新参数,在元测试中使更新模型与教师模型保持一致性损失。然而该方法要求对批次内的数据集标签找出最近的样本标签进行替换,难以满足网络流量数据集样本标签分布不均的实际性需求。文献[14]提出一种元软标签生成框架,使用干净数据为噪声数据生成软标签,然后在预测的软标签上进行训练,对数据集具有广泛的适用性。该方法需要在干净的训练集中进行元训练,使其分类器获得最佳性能,从而为其基分类器提供最耐噪声的学习。基于无噪声模型方法对于复杂、结构化的噪声效果并不明显,因为其无法处理一些特殊情况下的噪声标签。

在损失函数方面,针对带噪声的数据集的深度神经网络往往用交叉熵作为损失函数,然而交叉熵的类别偏向性会导致在整个训练中简单类比困难类更容易学习,且收敛得更快。这会导致当简单类已经过拟合于噪声标签时,困难类仍然存在学习的不足。

针对以上研究中出现的问题,本文提出了基于分层残差连接的深度残差神经网络(SCE-Res2Net, Symmetric Cross Entropy-Res2Net)方法对不同程度带噪声的数据集进行学习。该方法通过更小维度的残差块替代原本残差块进行分组卷积的方式,隐性地从训练数据里提取特征,并且不会随着网络层数的增加导致更大误差。同时通过对称学习的方式对交叉熵进行改进,增强了对困难类的学习且不会过拟合于简单类,解决了交叉熵的困难分类学习问题以及对噪声标签的过拟合问题,具有很好的鲁棒性;亦具有很好的抗噪性,在高噪声下具备较高的精度。该方法解决了网络流量分类需要大量干净标签的问题,提高了噪声容忍的效果,为实际大规模的真实世界网络噪声流量分类奠定了基础。

3 基于改进的深度残差神经网络的网络流量噪声容忍方法

采用 SCE-Res2Net 对网络流量数据集进行噪声容忍学习的流程包括:数据集的采集,数据集的预处理,模拟实际情况进行不同程度的加噪,特征提取及分类。其核心思想是通过引入带有更小维度滤波器的分层残差框架对数据集进行学习;同时通过对称学习的方法来解决交叉熵的困难标签分类问题以及噪声标签过拟合问题,从而完成对带有标签噪声的网络流量的噪声容忍。其整体架构如图 1 所示。

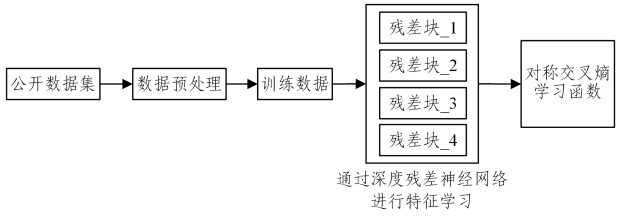

图 1 网络流量噪声容忍的整体架构

Fig. 1 Overall architecture of network traffic noise tolerance

3.1 深度残差网络结构

深度残差网络[15]与传统卷积神经网络[16]不同,其通过在卷积层中每隔一定层数构建一个残差块的方法进行残差学习,用来解决随着网络层数深度增加导致的梯度消散以及性能退化的问题。构建的残差块如图 2 所示。

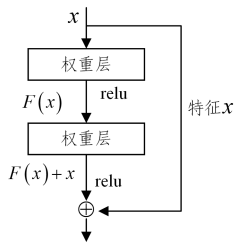

图 2 残差学习的构造块

Fig. 2 Building block of residual learning

其中残差学习的形式为

$$y=F(x,\{W_i\})+x \tag{1}$$

其中,x 和 y 是所考虑的层的输入和输出向量。函数 $F(x,\{W_i\})$ 表示需要学习的残差映射。图 2 中有两层卷积层时可以表示为:

$$F=W_2\sigma(W_1x) \tag{2}$$

其中,σ 表示 ReLU[17],为了简化符号,省略了偏差。图中 $F+x$ 是通过一个快捷连接和元素加法来执行的。本文采用加入后的第二次非线性结果,即 $\sigma(y)$。

x 和 F 的维数在式(1)中必须相等。当出现因改变输入/输出通道导致 x 和 F 维数不相等时,可以通过快捷连接进行线性投影 W_s 来匹配尺寸:

$$y=F(x,\{W_i\})+W_sx \tag{3}$$

也可以在式(1)中使用 W_s。由于标识映射足够解决退化问题并且具有较高的资源利用率,因此 W_s 只在匹配维数时使用。不同的 ResNet 深度残差神经网络的结构如表 1 所列。

表 1　5 种不同的 ResNet 深度残差神经网络结构

Table 1　5 different ResNet structures

层名称	输出大小	ResNet18	ResNet34	ResNet50	ResNet101	ResNet152
卷积层 1	112×112	7×7，64，stride 2				
		3×3 最大池化，stride 2				
卷积层 2	56×56	$\begin{bmatrix}3\times3, & 64\\ 3\times3, & 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 64\\ 3\times3, & 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\ 3\times3, & 64\\ 1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\ 3\times3, & 64\\ 1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\ 3\times3, & 64\\ 1\times1, & 256\end{bmatrix}\times3$
卷积层 3	28×28	$\begin{bmatrix}3\times3, & 128\\ 3\times3, & 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 128\\ 3\times3, & 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\ 3\times3, & 128\\ 1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\ 3\times3, & 128\\ 1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\ 3\times3, & 128\\ 1\times1, & 512\end{bmatrix}\times8$
卷积层 4	14×14	$\begin{bmatrix}3\times3, & 256\\ 3\times3, & 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 256\\ 3\times3, & 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\\ 1\times1, & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\\ 1\times1, & 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\\ 1\times1, & 1024\end{bmatrix}\times36$
卷积层 5	7×7	$\begin{bmatrix}3\times3, & 512\\ 3\times3, & 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 512\\ 3\times3, & 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\end{bmatrix}\times3$
	1×1	平均池化，1 000-d fc，softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

3.2　对称交叉熵学习结构

对称交叉熵学习方法 SCE[18] 是 2019 年设计的用于解决交叉熵在作为分类损失函数中存在的硬学习和标签噪声过拟合问题的方法。与目前最先进的方法相比，SCE 具有更好的鲁棒性，并且容易应用于现有的神经网络结构中。

对于 K 类数据集 $D=\{(x,y)^{(i)}\}_{i=1}^{n}$，传统的交叉熵损失函数可以如下表示：

$$l_{ce}=-\sum_{k=1}^{K}q(k|x)\log p(k|x) \tag{4}$$

其中，$x\in\chi\subset\mathbb{R}^d$ 为 d 维输入空间中的一个样本，$y\in Y=\{1,\cdots,K\}$ 是 x 的标签。对于每个样本 x，分类器 $f(x)$ 计算每个标签 $k\in\{1,\cdots,K\}$ 的概率为：

$$p(k|x)=\frac{e^{z_k}}{\sum_{j=1}^{K}e^{z_j}} \tag{5}$$

其中，e^{z_j} 为对数。$q(k|x)$ 表示样本 x 的真实标签分布并且 $\sum_{k=1}^{K}q(k|x)=1$。考虑单一真实标签值的标签 y 的情况，则对于所有 $k\neq y$ 时，$q(y|x)=1$，$q(k|x)=0$。

对于已知的两个分布 q 和 p，它们之间交叉熵的关系 $H(q\parallel p)$ 和 KL 散度 $KL(q\parallel p)$ 可以表示为：

$$KL(q\parallel p)=H(q\parallel p)-H(q) \tag{6}$$

其中，$H(q)$ 是 q 的熵，$q=q(k|x)$ 是以样本 x 为条件的真值类别分布，$p=p(k|x)$ 是通过分类器 f 在标签上的预测分布。从 KL 散度的角度来看，分类是学习一个接近真值类别分布 $q=q(k|x)$ 的预测分布 $p=p(k|x)$，目的是最小化两个分布之间的 KL 散度 $KL(q\parallel p)$。通常 $H(q(k|x))$ 在给定的类分布下是一个常数，因此式(6)中省略了式(4)中的交叉熵损失。

对于给定一个真实分布 q 和它的近似分布 p，$KL(q\parallel p)$ 使用针对 p 优化的代码(需要额外比特数的惩罚)对来自 q 的样本进行惩罚。在有噪声标签的情况下，我们知道 $q(k|x)$ 并不代表真实的类分布，而 $p(k|x)$ 在一定程度上可以反映真实的类分布。因此，除了把 $q(k|x)$ 作为真值外，还需要考虑 KL 散度的另一个方向，即 $KL(p\parallel q)$，在使用 $q(k|x)$ 的编码时，惩罚来自 $p(k|x)$ 的编码样本。对称 KL 散度为：

$$SKL=KL(q\parallel p)+KL(p\parallel q) \tag{7}$$

将这对称思想从 KL 散度传递到交叉熵，得到了对称交叉熵(SCE)为：

$$SCE=CE+RCE=H(q,p)+H(p,q) \tag{8}$$

其中，$RCE=H(p,q)$ 是 $H(q,p)$ 的反式，即反向交叉熵。样本 x 的 RCE 损失为：

$$l_{rce}=-\sum_{k=1}^{K}p(k|x)\log q(k|x) \tag{9}$$

每个样本的 SCE 损失为：

$$l_{sce}=l_{ce}+l_{rce} \tag{10}$$

虽然 RCE 项是耐噪声的，但 CE 项对于标签噪声不是鲁棒的。但 CE 的收敛性很好。因此采用灵活的对称学习框架 SL 以提高学习的有效性和鲁棒性。SL 损失可以表示为：

$$l_{sl}=\alpha l_{ce}+\beta l_{rce} \tag{11}$$

其中，α 和 β 为两个解耦的超参数，其中 α 用于研究 CE 的过拟合问题，β 用于灵活研究 RCE 的鲁棒性问题。

3.3　改进的深度残差神经网络结构

Res2Net 模型是文献[19]在 2021 年提出的一种多尺度网络结构的深度残差神经网络，用于解决目标检测、类激活映射以及显著目标检测等计算机视觉任务。Res2Net 通过使用更小的以残差层次化连接的滤波器组替换了 n 个通道的 3×3 卷积核，来增加输出特征所能代表的尺度的数量，同时保持相同的计算负荷，最终在更细粒度的程度上展现了多尺度表示能力。如图 3 所示，与传统的深度残差神经网络相比，经过 1×1 卷积后，Res2Net 将特征映射平均分割为 s 个特征映射子集，用 X_i 表示，其中 $i\in\{1,2,\cdots,s\}$。

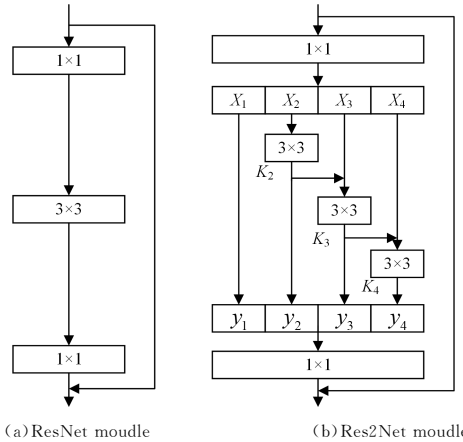

图 3　ResNet 与 Res2Net 瓶颈层对比

Fig. 3　Comparison in bottleneck block between ResNet and Res2Net

与输入特征图相比，每个特征子集 X_i 具有相同的空间大小，但通道数为 $1/s$。除 X_i 外，每个 X_i 都有对应的 3×3 卷积，用 $K_i(\ )$ 表示。用 y_i 表示 $K_i(\ )$ 的输出。将特征子集 X_i 与 $K_{i-1}(\ )$ 的输出相加，然后输入 $K_i(\ )$。为了在增加 s 的同时减少参数，省略了 X_i 的 3×3 卷积。因此，y_i 可以表示为：

$$y_i=\begin{cases}x_i, & i=1\\ K_i(x_i), & i=2\\ K_i(x_i+y_{i-1}), & 2<i\leqslant s\end{cases} \tag{12}$$

其中，每个 3×3 卷积操作 $K_i(\ )$ 都有可能从所有的特征分割 $\{x_j, j\leqslant i\}$ 中接收到特征信息。每次特征分割 x_j 经过 3 × 3 卷积操作时，输出结果的接收域都比 x_j 大。由于组合爆炸效应，Res2Net 模块的输出包含不同数量和不同组合的接收域大小/尺度。

我们使用过滤器宽度 w、尺寸 s 作为控制参数。更大的 s 允许具有更丰富的接受域大小的特征被学习，而通过连接引入的计算/内存消耗可以忽略不计。不同的 Res2net 深度残差神经网络的结构如表 2 所列。

表 2 5 种不同的 Res2Net 深度残差神经网络结构

Table 2 5 different Res2Net structures

类别	名称	滤波器宽度 w	尺度尺寸 s
1	Res2Net50_26w_4s	26	4
2	Res2Net50_48w_2s	48	2
3	Res2Net50_14w_8s	14	8
4	Res2Net50_26w_6s	26	6
5	Res2Net50_26w_8s	26	8
6	Res2Net101_26w_4s	26	4

4 实验结果

4.1 网络流量数据集的构建

带有标签噪声的数据集是保证本文模型学习性能的一个关键因素。深度残差神经网络的输入一般为二维矩阵，因此本节首先对需要处理的原始网络流量数据及标签的标注工作进行简单介绍。为了验证后面所涉及的基于深度残差神经网络的噪声容忍方法的性能，根据文献[19]的研究，选用表 2 中模型 1 作为基准，并对传统的 Res2Net 模型进行了如下改进，其采集和标注工作如下。

(1)根据对数据集的分析得到 16×16 的灰度图片，网络流量数据的噪声容忍学习的目的是在带有高比例噪声的数据中学习得到更高的精度，其中包括 11 个不同的应用类型。为了使图像可以适应神经网络神经元数量并且不至于过快丢失边缘信息，将图像缩放至 224×224。

(2)针对网络流量、数据流量类别比例差距过大的问题，采用图像增强的方式对数据集进行了数据增强，使得各类别数据基本保持在同一个维度。

(3)根据网络流量数据特点，设计了多种具有不同结构的损失函数进行烧蚀研究。通过对不同结构的精度进行对比，得到最适合网络流量的损失函数。

4.2 Moore 数据集

Moore 数据集[20]是两个位于不同国家的两个完全不同的研究中心站点中超过 1 000 名研究人员、管理人员和技术支持人员通过千兆以太网连接到互联网的网络数据。数据集捕获站点边界到 Internet 的全双工流量。本文采用第 3 天的 130 623 个网络样本，包含 11 个网络流量类别，249 个特征，其中，最后一项属性是每条网络流相对应的类别。首先，将数据集归一化以后转换成灰度图片，并将灰度图片放大到适合二维矩阵的维度以完成灰度图片的映射；然后，针对数据集不平衡的问题进行数据增强；最后，对干净的标签进行不同比例的加噪，以模拟真实世界的网络流量标签，并应用于模型中进行评估。Moore 数据集统计信息如表 3 所列。

本文分为训练数据集、验证数据集和测试数据集 3 个部分，这 3 个数据集中每一种类别的比例与原流量保持一致。随机选取 5 000 条数据作为验证数据集，5 000 条作为测试数据集，其他为训练集。

表 3 Moore 流量数据统计信息

Table 3 Statistics of Moore traffic data

类别	数量	比例/%	应用来源
WWW	109 130	83.546	Web browsers, web applications
MAIL	1 479	1.132	IMAP, POP, SMTP
BULK	3 160	2.419	FTP, wget
ATTACK	29	0.022	Port scans, worms, viruses, sql injections
CHAT	185	0.142	MSN Messenger, Yahoo IM, Jabber
P2P	10 871	8.322	Napster, Kazaa, Gnutella, eDonkey, BitTorrent
DATDABASE	5 387	4.124	MySQL, dbase, Oracle
MULTIMEDIA	73	0.056	Windows Media Player, Real, iTunes
VOIP	34	0.026	Skype
SERVICES	51	0.039	X11, DNS, IDENT, LDAP, NTP
INTERACTIVE	173	0.132	SSH, TELNET, VNC, GotoMyPC
?	51	0.039	未知
总计	130 623 000	100.000	

4.3 数据预处理

本文根据深度残差神经网络能够通过残差学习促进梯度传播的方式以使用更深层次模型的特点，对数据集进行如下预处理。

深度残差网络所需的数据集的二维矩阵可以表示如下：

$$\boldsymbol{T}=\begin{bmatrix}A_{11} & A_{12} & \cdots & A_{1m}\\ A_{21} & A_{22} & \cdots & A\\ \vdots & \vdots & \ddots & \vdots\\ A_{n1} & A_{n2} & \cdots & A_{mn}\end{bmatrix} \tag{13}$$

其中数据集可以表示为：

$$\boldsymbol{B}_i=[A_{1i},A_{2i},A_{3i},\cdots,A_{mi}]^{\mathrm{T}} \tag{14}$$

其中，$\boldsymbol{B}_i$ 表示所有数据的第 i 个特征值，那么矩阵 $\boldsymbol{T}$ 可以表示为：

$$\boldsymbol{T}=[\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3,\cdots,\boldsymbol{B}_m] \tag{15}$$

对每一列进行归一化处理后可以得到 B_i'：

$$\boldsymbol{B}_i'=\frac{\boldsymbol{B}_i}{\max\{A_{1i},A_{2i},A_{3i},\cdots,A_{mi}\}} \tag{16}$$

那么量化后的矩阵可以表示为：

$$\boldsymbol{T}_i'=[B_1^i,B_2^i,B_3^i,\cdots,B_m^i] \tag{17}$$

根据 Moore 数据集具有的 249 位特征，构建一个 16×16 的矩阵。由于最后一位特征为真实标签，于是对矩阵末位进行 8 次补 0 操作。由于 Moore 数据集中存在 51 条无法确定类别的网络流量数据，因此本文在处理中不考虑对其标注标签并加入模型训练。Moore 数据集中存在缺失值与 bool 型值，根据文献[21]中的预处理综述，将 65－68、71－72 的统计特征中‘Y’设置为 1，‘N’设置为 0，‘N/Y’设置为 3，缺失值为 4，将 68－70 统计特征中的缺失值设置为 999，表明这是一个新的特征。具体统计特征缺失值的填充如表 4 所列。

表 4　特征缺失值填充方式

Table 4　Feature missing value filling method

特征序号	填充值
63	1.0
64	1.0
65－68	3
69－70	999
71－72	2
73	0.0
74	0.0
209	0.18

将填充后归一化的矩阵中每个元素作为像素点映射到灰度图像中，每张灰色图像表示一条网络流量数据。图 4 为 Moore 数据集中比例最高的 4 类应用类型对应的灰度图片。

为了防止由于数据集类别及其不均衡可能导致的模型过拟合于最高比例类别，以及小比例类别数据难以识别的问题，本文对数据集进行了数据增强。采用图像增强领域常用的旋转、镜像、裁切、平移等方式对小比例类别流量图片数据进行了数据增强。所使用的最终增强后的网络流量数据集统计信息如表 5 所列。

(a)WWW　(b)P2P　(c)DATABASE　(d)BULK

图 4　不同 Moore 流量数据对应的灰度图片

Fig. 4　Gray scale images of different Moore traffic data

表 5　实际使用的 Moore 流量数据统计信息

Table 5　Statistics of Moore traffic data actually used

类别	训练集数量	验证集/测试集数量	比例/%
WWW	9891	511	10.24
MAIL	10724	554	11.10
BULK	8590	445	8.89
ATTACK	6306	327	6.53
CHAT	6704	348	6.94
P2P	9853	509	10.20
DATDABASE	9764	505	10.11
MULTIMEDIA	10584	548	10.96
VOIP	7394	383	7.65
SERVICES	7394	383	7.65
INTERACTIVE	9406	487	9.74
总计	96610	5000	

为了模拟真实世界网络流量分类标记不准确的情况，通过置换标签的方式对数据集添加对称标签噪声。标签为 i 的流量数据置换成标签 j 的数量可以表示为：

$$N_i(j)=N\times P_i\times\rho\times\left(\frac{P_j}{1-P_i}\right) \tag{18}$$

其中，N 为标签总数，ρ 为噪声比，P_i 为第 i 个标签在所有标签中的占比，P_j 为第 j 个标签在所有标签中的占比。本文在 20%～80%标签噪声的环境下进行实验，置换后的噪声标签数量如表 6 所列。

表 6　所需置换的噪声标签数量

Table 6　Number of noisy labels to be replaced

噪声比	0	1	2	3	4	5	6	7	8	9	10
0.2	1971	2138	1714	1256	1335	1965	1948	2111	1473	1473	1877
0.3	2961	3212	2571	1885	2007	2949	2924	3170	2211	2211	2814
0.4	3951	4283	3452	2517	2676	3936	3900	4226	2952	2952	3757
0.5	4939	5357	4289	3149	3347	4920	4875	5286	3691	3691	4698
0.6	5928	6428	5147	3778	4016	5905	5853	6345	4431	4431	5637
0.7	6918	7502	6009	4409	4687	6891	6828	7402	5169	5169	6580
0.8	7906	8574	6869	5038	5360	7877	7806	8461	5910	5910	7520

4.4　实验设置

本文实验环境和主要参数设置如表 7、表 8 所列。

表 7　实验环境参数

Table 7　Experimental environment parameters

类别	参数
硬件环境	Precision 5820 Tower X-Series
操作系统	Windows 10 64 bit
处理器	Intel(R) Core(TM) i9-10920X CPU @ 3.50 GHz 3.50 GHz
显卡	NVIDIA GeForce RTX 3080
内存/GB	32
Anaconda3	conda 4.10.1
Tensorflow	2.3.0 版本
CUDA	11.6

表 8　主要参数设置

Table 8　Main parameter settings

参数名称	参数值
learning rate(1－4epoch)	0.0008
learning rate(5－6epoch)	0.00008
Optimizer	SGD
epoch	6
batch	1000
batch_size	8
total training times	72

4.5　实验结果与分析

实验结果的评价为分类的整体准确率。为了寻找最优

模型，使用带有对称标签噪声的 Moore 数据集进行实验，并对损失函数进行烧蚀研究。在 0.7 的噪声比下的实验结果如表 9 所列。

表 9 使用不同 α，β 值的 SCE-Res2Net 的准确率

Table 9 Accuracy of SCE-Res2Net with different α and β values

（单位：%）

α	β				
	0.1	0.2	0.3	0.4	0.5
1	88.04	87.34	87.58	87.30	87.78
2	86.44	87.20	85.94	87.34	86.76
3	85.04	86.20	82.96	84.60	84.68
4	80.96	82.18	82.66	82.88	84.00

实验结果表明，在 0.7 的噪声下，我们模型的准确率普遍增加。当 $\alpha=1$，$\beta=0.1$ 时，模型的分类准确率最高，因此选取该模型为最优模型。同时将最优的模型与基线进行对比，实验结果如表 10 所列。

表 10 采用不同算法的准确率

Table 10 Accuracy of different algorithms

（单位：%）

算法	噪声比						
	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ResNet+CE	96.50	95.48	93.98	92.34	89.32	84.08	69.22
Res2net+CE	96.92	95.94	94.82	93.06	90.68	86.92	76.74
SCE-Res2net	97.26	96.02	95.00	93.38	91.00	89.84	79.48

上述实验结果说明，本文所提出的 SCE-Res2Net 算法，对 Moore 数据集进行预处理后转换成灰度图片，然后对其进行数据增强，模拟真实世界情况对其进行不同程度的加噪，相较于使用交叉熵作为损失函数的 ResNet 和 Res2Net，均有明显的提升，且在高噪声比下提升显著。在 0.7 噪声比下，本文算法相较于基线最高可以提升 2.92%。因此对称学习解决了带标签噪声的网络流量数据困难类的学习以及简单类的过拟合问题，并取得了很好的效果。SCE-Res2Net 方法在对噪声流量数据的噪声上的成功为接下来真实世界的噪声流量学习奠定了基础。

结束语 本文首先介绍了常用的噪声处理方法，然后基于相关研究，给出了一种神经网络损失函数的 SCE-Res2Net 算法，并将其应用于对网络流量数据的标签噪声学习过程，设计不同的损失函数完成最优分类模型的选取，使其在简单类不容易过拟合，解决了困难类的学习问题，从而提高了噪声容忍的效果。在未来的工作中，将针对以下 3 个方面进行更深一步的研究：(1)将模型应用到加密流量数据集中进行实验，探索加密流量噪声容忍的实用性；(2)利用大数据分析平台实现算法的分布式计算，提高噪声容忍的效率；(3)将元学习的方式应用到模型训练中，与当前模型的噪声分类效果进行比较。

参 考 文 献

[1] ERMAN J, MAHANTI A, ARLITT M. Qrp05-4: Internet traffic identification using machine learning[C]// IEEE Globecom 2006. IEEE, 2006: 1-6.

[2] WANG Y, ZHOU H Y, FENG H, et al. A Network Traffic Classification Method Based on Deep Convolution Neural Network[J]. Journal on Communications, 2018, 39(1): 14-23.

[3] DONG S, XIA Y, PENG T. Traffic identification model based on generative adversarial deep convolutional network[J]. Annals of Telecommunications, 2022, 77(9/10): 573-587.

[4] XUE W L, YU J, GUO Z Q, et al. End-to-end encrypted traffic classification based on a feature fusion convolutional neural network [J]. Computer Engineering and Application, 2021, 57(18): 8.

[5] TANAKA D, IKAMI D, YAMASAKI T, et al. Joint optimization framework for learning with noisy labels[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5552-5560.

[6] ZHANG C, BENGIO S, HARDT M, et al. Understanding deep learning(still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107-115.

[7] ARPIT D, JASTRZBSKI S, BALLAS N, et al. A closer look at memorization in deep networks[C]// International Conference on Machine Learning. PMLR, 2017: 233-242.

[8] XIAO T, XIA T, YANG Y, et al. Learning from massive noisy labeled data for image classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2691-2699.

[9] ZHENG S, WU P, GOSWAMI A, et al. Error-bounded correction of noisy labels[C]// International Conference on Machine Learning. PMLR, 2020: 11447-11457.

[10] HUANG J, QU L, JIA R, et al. O2u-net: A simple noisy label detection approach for deep neural networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3326-3334.

[11] SHARMA K, DONMEZ P, LUO E, et al. Noiserank: Unsupervised label noise reduction with dependence models[C]// European Conference on Computer Vision. Cham: Springer, 2020: 737-753.

[12] CHAROENPHAKDEE N, LEE J, SUGIYAMA M. On symmetric losses for learning from corrupted labels[C]// International Conference on Machine Learning. PMLR, 2019: 961-970.

[13] LI J, WONG Y, ZHAO Q, et al. Learning to learn from noisy labeled data[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5051-5059.

[14] ALGAN G, ULUSOY I. Meta soft label generation for noisy labels[C]// 2020 25th International Conference on Pattern Recognition(ICPR). IEEE, 2021: 7142-7148.

[15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.

[17] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]// ICML. 2010.

[18] WANG Y, MA X, CHEN Z, et al. Symmetric cross entropy for robust learning with noisy labels[C]// Proceedings of the IEEE/

CVF International Conference on Computer Vision. 2019：322-330.

[19] GAO S H，CHENG M M，ZHAO K，et al. Res2net：A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2019，43(2)：652-662.

[20] LI W，CANINI M，MOORE A W，et al. Efficient application identification and the temporal and spatial stability of classification schema[J]. Computer Networks，2009，53(6)：790-809.

[21] WANG F Y. Machine Learning in Network Traffic Classification[D]. Chengdu：University of Electronic Science and Technology of China，2023.

MA Jiye，born in 1997. His main research interests include future networks and so on.

ZHU Guosheng，born in 1972，Ph.D，professor，is a member of China Computer Federation. His main research interests include future networks and so on.