(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111340191 B
(45)授权公告日 2023.02.21

(21)申请号 202010122760.5
(22)申请日 2020.02.27
(65)同一申请的已公布的文献号
申请公布号 CN 111340191 A
(43)申请公布日 2020.06.26
(73)专利权人 福州大学
地址 350108 福建省福州市闽侯县福州大学城乌龙江北大道2号福州大学
(72)发明人 陈羽中 张毓东
(74)专利代理机构 福州元创专利商标代理有限公司 35100
专利代理师 丘鸿超 蔡学俊
(51)Int.Cl.
G06N 3/0464(2023.01)
G06N 3/047(2023.01)
G06N 3/084(2023.01)
G06N 20/20(2019.01)
H04L 9/40(2022.01)

(56)对比文件
CN 108881192 A,2018.11.23
CN 110765458 A,2020.02.07
US 2018150635 A1,2018.05.31
US 2017134404 A1,2017.05.11
蒋鸿玲等.基于神经网络的僵尸网络检测.《智能系统学报》.2013,第8卷(第02期),23-28.
王伟.基于深度学习的网络流量分类及异常检测方法研究.《中国博士学位论文全文数据库信息科技辑》.2018,I139-3.
Tangda Yu 等.An Encrypted Malicious Traffic Detection System Based on Neural Network.《2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery 》.2020,

审查员 吴俊杰

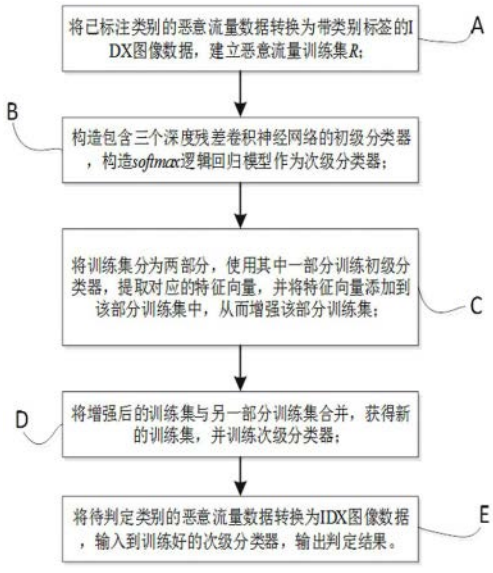
权利要求书3页 说明书6页 附图3页

(54)发明名称

基于集成学习的僵尸网络恶意流量分类方法及系统

(57)摘要

本发明涉及一种基于集成学习的僵尸网络恶意流量分类方法及系统，该方法包括：步骤A：将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据，建立恶意流量训练集R；步骤B：构造包含三个深度残差卷积神经网络的初级分类器，构造$softmax$逻辑回归模型作为次级分类器；步骤C：将训练集R分为R_0和R_1两部分，使用R_0训练初级分类器，提取恶意流量特征向量，并将提取的特征向量添加到R_0中，增强该部分训练集；步骤D：将增强后的训练集与R_1合并，用其训练次级分类器；步骤E：将待判定类别的恶意流量数据转换为IDX图像格式，输入到训练好的次级分类器，输出判定结果。该方法及系统有利于快速、准确地识别恶意流量类别。



CN 111340191 B

1.一种基于集成学习的僵尸网络恶意流量分类方法,其特征在于,包括以下步骤:

步骤A:将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据,建立恶意流量训练集R;

步骤B:构造包含三个深度残差卷积神经网络的初级分类器,构造softmax逻辑回归模型作为次级分类器;

步骤C:将训练集R分为R_0和R_1两部分,使用R_0训练初级分类器,提取恶意流量特征向量,并将提取的特征向量添加到R_0中,以增强该部分训练集;

步骤D:将增强后的训练集与R_1合并,用其训练次级分类器;

步骤E:将待判定类别的恶意流量数据转换为IDX图像格式,输入到训练好的次级分类器,输出判定结果;

所述步骤A具体包括以下步骤:

步骤A1:从已标注类别的恶意流量数据中清除没有应用层数据的数据报文;

步骤A2:对步骤A1处理后的恶意流量数据进行划分,将属于同一TCP会话的恶意流量数据划分为一组,将恶意流量数据中的网络层、传输层、应用层报文信息保存到一个二进制文件中;

步骤A3:将步骤A2得到的二进制文件截断或补0x00到固定长度M个字节,以保留TCP会话中能够反映流量类别特征的网络层首部、传输层首部、应用层首部信息以及部分应用层数据,去除不能反映流量类别特征的其他应用层数据;以字节为单位,将每个字节转换为灰度值,输出大小为m×m的灰度图像文件,$M=m^2$;

步骤A4:将步骤A3得到的灰度图像文件转换为IDX图像文件;

步骤A5:遍历已标注类别的恶意流量数据,得到训练集R;

其中$R=\{(x_i,y_i)\mid i=1,2,...,N\}$,N为训练集R中的恶意流量样本数,$x_i$为一个IDX图像文件,$y_i$为对应的恶意流量类别标签;$y_i\in C=\{1,2,...,K\}$,C表示恶意流量类别标签集合,$y_i=c$,$1\leqslant c\leqslant K$,表示$x_i$为第c种恶意流量类别。

2.根据权利要求1所述的基于集成学习的僵尸网络恶意流量分类方法,其特征在于,所述恶意流量类别为包括Cridex、Geodo、Htbot、Miuref、Neris、Nsisay、Shifu、Virut和Zeus的僵尸网络恶意流量。

3.根据权利要求1所述的基于集成学习的僵尸网络恶意流量分类方法,其特征在于,所述步骤B中,所述三个深度残差卷积神经网络分别为$ResNet_1$、$ResNet_2$和$ResNet_3$;每个深度残差卷积神经网络包括五个残差单元和两个全连接层;其中,每个残差单元按照输入样本的数据流向依次包括1个卷积层、1个批量正则化层、1个线性激活层和2个卷积层。

4.根据权利要求3所述的基于集成学习的僵尸网络恶意流量分类方法,其特征在于,所述步骤C具体包括以下步骤:

步骤C1:将数据集R随机分为R_0和R_1两个训练子集,再将R_0随机拆分成三个训练子集$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$;

步骤C2:分别使用$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$三个训练子集对三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$进行训练;

步骤C3:利用步骤C2训练好的三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$

分别对R_0中的每个IDX图像样本进行恶意流量特征向量提取，然后对$ResNet_1$、$ResNet_2$和$ResNet_3$获得的特征向量求平均，输出与各IDX图像样本对应的特征向量，并将各特征向量覆盖到对应的IDX图像样本的二进制文件末尾，保持文件字节数不变；遍历R_0中的每个IDX图像样本后，得到特征增强后的训练子集R_0^+。

5.根据权利要求4所述的基于集成学习的僵尸网络恶意流量分类方法，其特征在于，所述步骤C2中，使用训练子集$R_0^{(i)}$，采用均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，采用交叉熵作为损失函数计算损失值，利用反向传播迭代更新模型参数，以最小化损失函数作为训练目标，对相应的深度残差卷积神经网络$ResNet_i$进行训练。

6.根据权利要求4所述的基于集成学习的僵尸网络恶意流量分类方法，其特征在于，所述步骤D具体包括以下步骤：

步骤D1：合并训练子集R_1和步骤C3得到的训练子集R_0^+，得到增强后的训练集R^+；

步骤D2：使用R^+训练次分类器的softmax逻辑回归模型，用交叉熵作为损失函数计算损失值，通过均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，利用反向传播迭代更新模型参数，以最小化损失函数来训练模型，得到训练好的softmax逻辑回归模型。

7.根据权利要求1所述的基于集成学习的僵尸网络恶意流量分类方法，其特征在于，所述步骤E具体包括以下步骤：

步骤E1：按照步骤A1-A4，将待判定类别的僵尸网络恶意流量数据转换为IDX图像格式，表示为m×m的二维向量矩阵X，计算矩阵X的协方差矩阵cov；

$$cov=\frac{1}{m-1}X^TX$$

步骤E2：计算协方差矩阵cov的特征值与特征向量，根据特征值大小对特征向量排序，保留前K个特征向量，对前K个特征向量进行平均，得到特征向量$\vec{x}$，其中K是步骤A5中所述恶意流量类别标签集合C中的类别标签数；

步骤E3：将$\vec{x}$输入到DropOut层，然后输入训练好的softmax逻辑回归模型，计算该流量数据属于流量类别c的概率$p(y_i=c)\big|_{c=1}^{K}$，选择$\underset{c\in C}{argmax}\,p(y_i=c)$作为判定的僵尸网络恶意流量类别，输出判定结果。

8.一种采用如权利要求1-7任一项所述方法的基于集成学习的僵尸网络恶意流量分类系统，其特征在于，包括：

数据收集模块，用于根据TCP连接作为标准划分网络流量，以产生流量的僵尸网络种类作为恶意流量种类，收集恶意流量数据并标注类别；

数据预处理模块，用于将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据，构建恶意流量训练集R，并将其分为R_0和R_1两个训练子集，对其中的R_0进行数据增强；

数据增强模块，用于构造包含三个深度残差卷积神经网络的初级分类器和构造softmax逻辑回归模型作为次级分类器，然后使用R_0训练初级分类器，提取恶意流量特征向量，并用提取的特征向量增强训练子集R_0；

次级分类器训练模块，用于将增强后的训练子集R_0和训练子集R_1合并，并用合并后的训

练集训练次级分类器;以及

恶意流量种类预测模块,用于利用训练好的次级分类器对输入的待判定类别的恶意流量数据进行预测,输出其所属的类别。

基于集成学习的僵尸网络恶意流量分类方法及系统

技术领域

[0001] 本发明属于网络安全领域,具体涉及一种基于集成学习的僵尸网络恶意流量分类方法及系统。

背景技术

[0002] 恶意流量特征提取一直是网络安全领域的难点问题。恶意软件可利用伪装、加密、欺骗、零日漏洞等技术实现行为的深度隐藏且它们可以频繁地变种,这些致使互联网中大量的僵尸网络恶意流量未被发现。由于僵尸主机产生的流量与正常主机产生的网络流量在特征上与有很大差异,通过对网络流量分类识别僵尸网络恶意流量也是检测僵尸网络的主要方向。因此对僵尸网络恶意流量分类投入研究有很重要的意义。

[0003] 目前有很多种网络流量异常检测方法,如基于统计、聚类、分类、信息熵等等。其中,将网络流量归类至特定的类型是其中很重要的一个方向,从而区分正常和僵尸网络恶意流量,并识别僵尸网络恶意流量类型。网络流量异常检测作为一种有效的网络防护手段,能够检测未知攻击行为,并为网络态势感知提供重要支持,按照使用技术的不同,目前一般的网络流量分类方法可以分为四类:基于端口识别的方法,基于深层包检测的方法,基于统计的方法,以及基于行为的方法。迄今为止,国内外学者基于这四类方向已经提出了很多不同类型的检测方法。但是,目前大多数网络流量分类方法都是基于传统的机器学习方式,分类性能非常依赖于流量特征的设计。

发明内容

[0004] 本发明的目的在于提供一种基于集成学习的僵尸网络恶意流量分类方法及系统,该方法及系统有利于快速、准确地识别恶意流量类别。

[0005] 为实现上述目的,本发明采用的技术方案是:一种基于集成学习的僵尸网络恶意流量分类方法,包括以下步骤:

[0006] 步骤A:将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据,建立恶意流量训练集R;

[0007] 步骤B:构造包含三个深度残差卷积神经网络的初级分类器,构造softmax逻辑回归模型作为次级分类器;

[0008] 步骤C:将训练集R分为R_0和R_1两部分,使用R_0训练初级分类器,提取恶意流量特征向量,并将提取的特征向量添加到R_0中,以增强该部分训练集;

[0009] 步骤D:将增强后的训练集与R_1合并,用其训练次级分类器;

[0010] 步骤E:将待判定类别的恶意流量数据转换为IDX图像格式,输入到训练好的次级分类器,输出判定结果。

[0011] 进一步地,所述步骤A具体包括以下步骤:

[0012] 步骤A1:从已标注类别的恶意流量数据中清除没有应用层数据的数据报文;

[0013] 步骤A2:对步骤A1处理后的恶意流量数据进行划分,将属于同一TCP会话的恶意流

量数据划分为一组，将恶意流量数据中的网络层、传输层、应用层报文信息保存到一个二进制文件中；

[0014] 步骤A3：将步骤A2得到的二进制文件截断或补0x00到固定长度M个字节，以保留TCP会话中能够反映流量类别特征的网络层首部、传输层首部、应用层首部信息以及部分应用层数据，去除不能反映流量类别特征的其他应用层数据；以字节为单位，将每个字节转换为灰度值，输出大小为m×m的灰度图像文件，$M=m^2$；

[0015] 步骤A4：将步骤A3得到的灰度图像文件转换为IDX图像文件；

[0016] 步骤A5：遍历已标注类别的恶意流量数据，得到训练集R；

[0017] 其中$R=\{(x_i,y_i)\mid i=1,2,...,N\}$，N为训练集R中的恶意流量样本数，$x_i$为一个IDX图像文件，$y_i$为对应的恶意流量类别标签；$y_i\in C=\{1,2,...,K\}$，C表示恶意流量类别标签集合，$y_i=c$，$1\leqslant c\leqslant K$，表示$x_i$为第c种恶意流量类别。

[0018] 进一步地，所述恶意流量类别为包括Cridex、Geodo、Htbot、Miuref、Neris、Nsisay、Shifu、Virut和Zeus的僵尸网络恶意流量。

[0019] 进一步地，所述步骤B中，所述三个深度残差卷积神经网络分别为$ResNet_1$、$ResNet_2$和$ResNet_3$；每个深度残差卷积神经网络包括五个残差单元和两个全连接层；其中，每个残差单元按照输入样本的数据流向依次包括1个卷积层、1个批量正则化层、1个线性激活层和2个卷积层。

[0020] 进一步地，所述步骤C具体包括以下步骤：

[0021] 步骤C1：将数据集R随机分为R_0和R_1两个训练子集，再将R_0随机拆分成三个训练子集$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$；

[0022] 步骤C2：分别使用$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$三个训练子集对三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$进行训练；

[0023] 步骤C2：分别使用$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$三个训练子集对三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$进行训练；

[0024] 步骤C3：利用步骤C2训练好的三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$分别对R_0中的每个IDX图像样本进行恶意流量特征向量提取，然后对$ResNet_1$、$ResNet_2$和$ResNet_3$获得的特征向量求平均，输出与各IDX图像样本对应的特征向量，并将各特征向量覆盖到对应的IDX图像样本的二进制文件末尾，保持文件字节数不变；遍历R_0中的每个IDX图像样本后，得到特征增强后的训练子集R_0^+。

[0025] 进一步地，所述步骤C2中，使用训练子集$R_0^{(i)}$，采用均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，采用交叉熵作为损失函数计算损失值，利用反向传播迭代更新模型参数，以最小化损失函数作为训练目标，对相应的深度残差卷积神经网络$ResNet_i$进行训练。

[0026] 进一步地，所述步骤D具体包括以下步骤：

[0027] 步骤D1：合并训练子集R_1和步骤C3得到的训练子集R_0^+，得到增强后的训练集R^+；

[0028] 步骤D2：使用R^+训练次分类器的softmax逻辑回归模型，用交叉熵作为损失函数计算损失值，通过均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，利

用反向传播迭代更新模型参数，以最小化损失函数来训练模型，得到训练好的softmax逻辑回归模型。

[0029] 进一步地，所述步骤E具体包括以下步骤：

[0030] 步骤E1：按照步骤A1-A4，将待判定类别的僵尸网络恶意流量数据转换为IDX图像格式，表示为m×m的二维向量矩阵X，计算矩阵X的协方差矩阵cov；

[0031] $$cov = \frac{1}{m-1} X^T X$$

[0032] 步骤E2：计算协方差矩阵cov的特征值与特征向量，根据特征值大小对特征向量排序，保留前K个特征向量，对前K个特征向量进行平均，得到特征向量$\vec{x}$，其中K是步骤A5中所述恶意流量类别标签集合C中的类别标签数；

[0033] 步骤E3：将$\vec{x}$输入到DropOut层，然后输入训练好的softmax逻辑回归模型，计算该流量数据属于流量类别c的概率$p(y_i = c)\Big|_{c=1}^{K}$，选择$\underset{c \in C}{argmax}\, p(y_i = c)$作为判定的僵尸网络恶意流量类别，输出判定结果。

[0034] 本发明还提供了一种基于集成学习的僵尸网络恶意流量分类系统，包括：

[0035] 数据收集模块，用于根据TCP连接作为标准划分网络流量，以产生流量的僵尸网络种类作为恶意流量种类，收集恶意流量数据并标注类别；

[0036] 数据预处理模块，用于将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据，构建恶意流量训练集R，并将其分为R_0和R_1两个训练子集，对其中的R_0进行数据增强；

[0037] 数据增强模块，用于构造包含三个深度残差卷积神经网络的初级分类器和构造softmax逻辑回归模型作为次级分类器，然后使用R_0训练初级分类器，提取恶意流量特征向量，并用提取的特征向量增强训练子集R_0；

[0038] 次级分类器训练模块，用于将增强后的训练子集R_0和训练子集R_1合并，并用合并后的训练集训练次级分类器；以及

[0039] 恶意流量种类预测模块，用于利用训练好的次级分类器对输入的待判定类别的恶意流量数据进行预测，输出其所属的类别。

[0040] 相较于现有技术，本发明具有以下有益效果：提供了一种基于集成学习的僵尸网络恶意流量分类方法及系统，首先将流量转换为图像的方式，利用图像识别领域表现优秀的模型和集成思想，使用残差网络作为初级分类器提取流量特征，利用神经网络提取流量的特征，不需要像传统机器学习一样设计能准确反映流量特征的特征集，也无需借助其解析信息和端口信息，从而解决现实中僵尸网络恶意流量大多被加密的问题；之后使用神经网络所提取的流量特征向量增强原训练集，用于训练网络参数较少、分类速度更快的基于Softmax逻辑回归的次级分类器，提升次级分类器的恶意流量分类能力，在识别精度和流量分类速度之间达到了较好的平衡。因此，本发明可进一步提升恶意流量识别性能，具有很强的实用性和广阔的应用前景。

附图说明

[0041] 图1是本发明实施例的方法实现流程图。

[0042] 图2是本发明实施例中步骤A的实现流程图。

[0043] 图3是本发明实施例的系统结构示意图。

具体实施方式

[0044] 下面结合附图及具体实施例对本发明作进一步的详细说明。

[0045] 本发明提供一种基于集成学习的僵尸网络恶意流量分类方法，如图1所示，包括以下步骤：

[0046] 步骤A：将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据，建立恶意流量训练集R。如图2所示，步骤A具体包括以下步骤：

[0047] 步骤A1：从已标注类别的恶意流量数据中清除没有应用层数据的数据报文。

[0048] 步骤A2：对步骤A1处理后的恶意流量数据进行划分，将属于同一TCP会话的恶意流量数据划分为一组，将恶意流量数据中的网络层、传输层、应用层报文信息保存到一个二进制文件中。

[0049] 其中，同一TCP会话的数据报文具有相同的五元组，即源IP地址、目的IP地址、目的端口、源端口和传输层协议。其中，(源IP地址:源端口)和(目的IP地址:目的端口)可以互换，TCP会话包含双向TCP数据流。

[0050] 步骤A3：将步骤A2得到的二进制文件截断或补0x00到固定长度M个字节，其目的是保留TCP会话中能够反映流量类别特征的网络层首部、传输层首部、应用层首部信息以及部分应用层数据，去除不能反映流量类别特征的其他大部分应用层数据。以字节为单位，将每个字节转换为灰度值，输出大小为m×m的灰度图像文件，其中$M=m^2$。一般可取M为784字节，m＝28，图像文件大小为28×28。

[0051] 步骤A4：将步骤A3得到的灰度图像文件转换为IDX图像文件。

[0052] 其中，IDX文件是索引文件格式，包含图片的像素及统计信息，是深度学习模型常用的输入图像数据格式。

[0053] 步骤A5：遍历已标注类别的恶意流量数据，得到训练集R。

[0054] 其中，$R=\{(x_i,y_i)\mid i=1,2,...,N\}$，N为训练集R中的恶意流量样本数，$x_i$为一个IDX图像文件，$y_i$为对应的恶意流量类别标签；$y_i\in C=\{1,2,...,K\}$，C表示恶意流量类别标签集合，恶意流量类别包括Cridex、Geodo、Htbot、Miuref、Neris、Nsisay、Shifu、Virut和Zeus等僵尸网络恶意流量，$y_i=c,1\leqslant c\leqslant K$，表示$x_i$为第c种恶意流量类别。

[0055] 步骤B：构造包含三个深度残差卷积神经网络的初级分类器，构造softmax逻辑回归模型作为次级分类器。

[0056] 其中，所述三个深度残差卷积神经网络分别为$ResNet_1$、$ResNet_2$和$ResNet_3$；每个深度残差卷积神经网络包括五个残差单元和两个全连接层。

[0057] 其中，每个残差单元按照输入样本的数据流向依次包括1个卷积层、1个批量正则化层、1个线性激活层和2个卷积层。

[0058] 第一残差单元的组成按照输入样本的数据流向依次为卷积层1，参数包括通道数为128，卷积核尺寸为3和3，步长为2；批量正则化层1；线性激活层1；卷积层2，参数包括通道数为256，卷积核尺寸为1和1，步长为1；批量正则化层2；线性激活层2。

[0059] 第二残差单元的组成按照输入样本的数据流向依次为卷积层1，参数包括通道数

为256，卷积核尺寸为3和3，步长为2；批量正则化层1；线性激活层1；卷积层2，参数包括通道数为512，卷积核尺寸为1和1，步长为1；批量正则化层2；线性激活层2。

[0060] 第三残差单元的组成按照输入样本的数据流向依次为卷积层1，参数包括通道数为512，卷积核尺寸为3和3，步长为2；批量正则化层1；线性激活层1；卷积层2，参数包括通道数为1024，卷积核尺寸为1和1，步长为1；批量正则化层2；线性激活层2。

[0061] 第四残差单元的组成按照输入样本的数据流向依次为卷积层1，参数包括通道数为1024，卷积核尺寸为3和3，步长为2；批量正则化层1；线性激活层1；卷积层2，参数包括通道数为2048，卷积核尺寸为1和1，步长为1；批量正则化层2；线性激活层2。

[0062] 第五残差单元的组成按照输入样本的数据流向依次为卷积层1，参数包括通道数为2048，卷积核尺寸为3和3，步长为2；批量正则化层1；线性激活层1；卷积层2，参数包括通道数为4096，卷积核尺寸为1和1，步长为1；批量正则化层2；线性激活层2。

[0063] 步骤C：将训练集R分为R_0和R_1两部分，使用R_0训练初级分类器，提取恶意流量特征向量，并将提取的特征向量添加到R_0中，以增强该部分训练集。具体包括以下步骤：

[0064] 步骤C1：将数据集R随机分为R_0和R_1两个训练子集，再将R_0随机拆分成三个，得到R_0的三个训练子集$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$。

[0065] 步骤C2：分别使用$R_0^{(1)}$、$R_0^{(2)}$和$R_0^{(3)}$三个训练子集对三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$进行训练。

[0066] 具体地，使用训练子集$R_0^{(i)}$，采用均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，采用交叉熵作为损失函数计算损失值，利用反向传播迭代更新模型参数，以最小化损失函数作为训练目标，对相应的深度残差卷积神经网络$ResNet_i$进行训练。

[0067] 步骤C3：利用步骤C2训练好的三个深度残差卷积神经网络$ResNet_1$、$ResNet_2$和$ResNet_3$分别对R_0中的每个IDX图像样本进行恶意流量特征向量提取，然后对$ResNet_1$、$ResNet_2$和$ResNet_3$获得的特征向量求平均，输出与各IDX图像样本对应的特征向量，并将各特征向量覆盖到对应的IDX图像样本的二进制文件末尾，保持文件字节数不变；遍历R_0中的每个IDX图像样本后，得到特征增强后的训练子集R_0^+。

[0068] 步骤D：将增强后的训练集与R_1合并，用其训练次级分类器。具体包括以下步骤：

[0069] 步骤D1：合并训练子集R_1和步骤C3得到的训练子集R_0^+，得到增强后的训练集R^+。

[0070] 步骤D2：使用R^+训练次分类器的softmax逻辑回归模型，用交叉熵作为损失函数计算损失值，通过均方根随机梯度下降优化方法RMsprop计算所有的梯度的平方的平均值，利用反向传播迭代更新模型参数，以最小化损失函数来训练模型，得到训练好的softmax逻辑回归模型。

[0071] 步骤E：将待判定类别的恶意流量数据转换为IDX图像格式，输入到训练好的次级分类器，输出判定结果。具体包括以下步骤：

[0072] 步骤E1：按照步骤A1-A4，将待判定类别的僵尸网络恶意流量数据转换为IDX图像格式，表示为m×m的二维向量矩阵X，计算矩阵X的协方差矩阵cov；

[0073] $$cov=\frac{1}{m-1}X^TX$$

[0074] 步骤E2:计算协方差矩阵cov的特征值与特征向量,根据特征值大小对特征向量排序,保留前K个特征向量,对前K个特征向量进行平均,得到特征向量$\vec{x}$,其中K是步骤A5中所述恶意流量类别标签集合C中的类别标签数;

[0075] 步骤E3:将$\vec{x}$输入到DropOut层(DropOut层是用于解决过拟合的一种机制,即对于神经网络单元,按照一定的概率将其暂时从网络中丢弃),然后输入训练好的softmax逻辑回归模型,计算该流量数据属于流量类别c的概率$p(y_i=c)\Big|_{c=1}^{K}$,选择$\underset{c\in C}{argmax}\, p(y_i=c)$作为判定的僵尸网络恶意流量类别,输出判定结果。

[0076] 本发明还提供了采用上述方法的基于集成学习的僵尸网络恶意流量分类系统,如图3所示,包括数据收集模块、数据预处理模块、数据增强模块、次级分类器训练模块和恶意流量种类预测模块。

[0077] 所述数据收集模块用于根据TCP连接作为标准划分网络流量,以产生流量的僵尸网络种类作为恶意流量种类,收集恶意流量数据并标注类别。

[0078] 所述数据预处理模块用于将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据,构建恶意流量训练集R,并将其分为R_0和R_1两个训练子集,对其中的R_0进行数据增强。

[0079] 所述数据增强模块用于构造包含三个深度残差卷积神经网络的初级分类器和构造softmax逻辑回归模型作为次级分类器,然后使用R_0训练初级分类器,提取恶意流量特征向量,并用提取的特征向量增强训练子集R_0。

[0080] 所述次级分类器训练模块用于将增强后的训练子集R_0和训练子集R_1合并,并用合并后的训练集训练次级分类器。

[0081] 所述恶意流量种类预测模块用于利用训练好的次级分类器对输入的待判定类别的恶意流量数据进行预测,输出其所属的类别。

[0082] 以上是本发明的较佳实施例,凡依本发明技术方案所作的改变,所产生的功能作用未超出本发明技术方案的范围时,均属于本发明的保护范围。

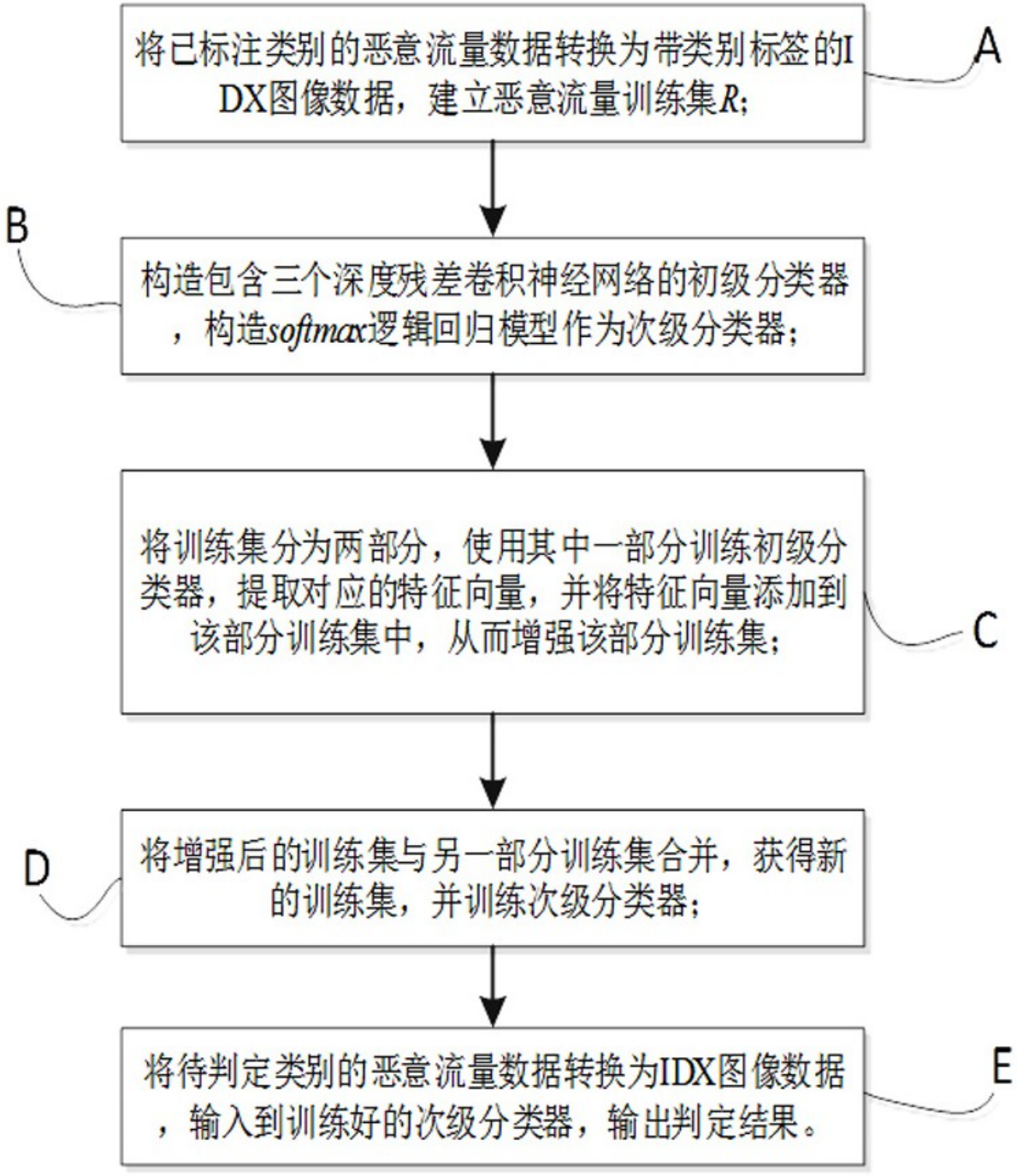

将已标注类别的恶意流量数据转换为带类别标签的IDX图像数据，建立恶意流量训练集R；
A
B
构造包含三个深度残差卷积神经网络的初级分类器，构造softmax逻辑回归模型作为次级分类器；
将训练集分为两部分，使用其中一部分训练初级分类器，提取对应的特征向量，并将特征向量添加到该部分训练集中，从而增强该部分训练集；
C
D
将增强后的训练集与另一部分训练集合并，获得新的训练集，并训练次级分类器；
将待判定类别的恶意流量数据转换为IDX图像数据，输入到训练好的次级分类器，输出判定结果。
E

图1

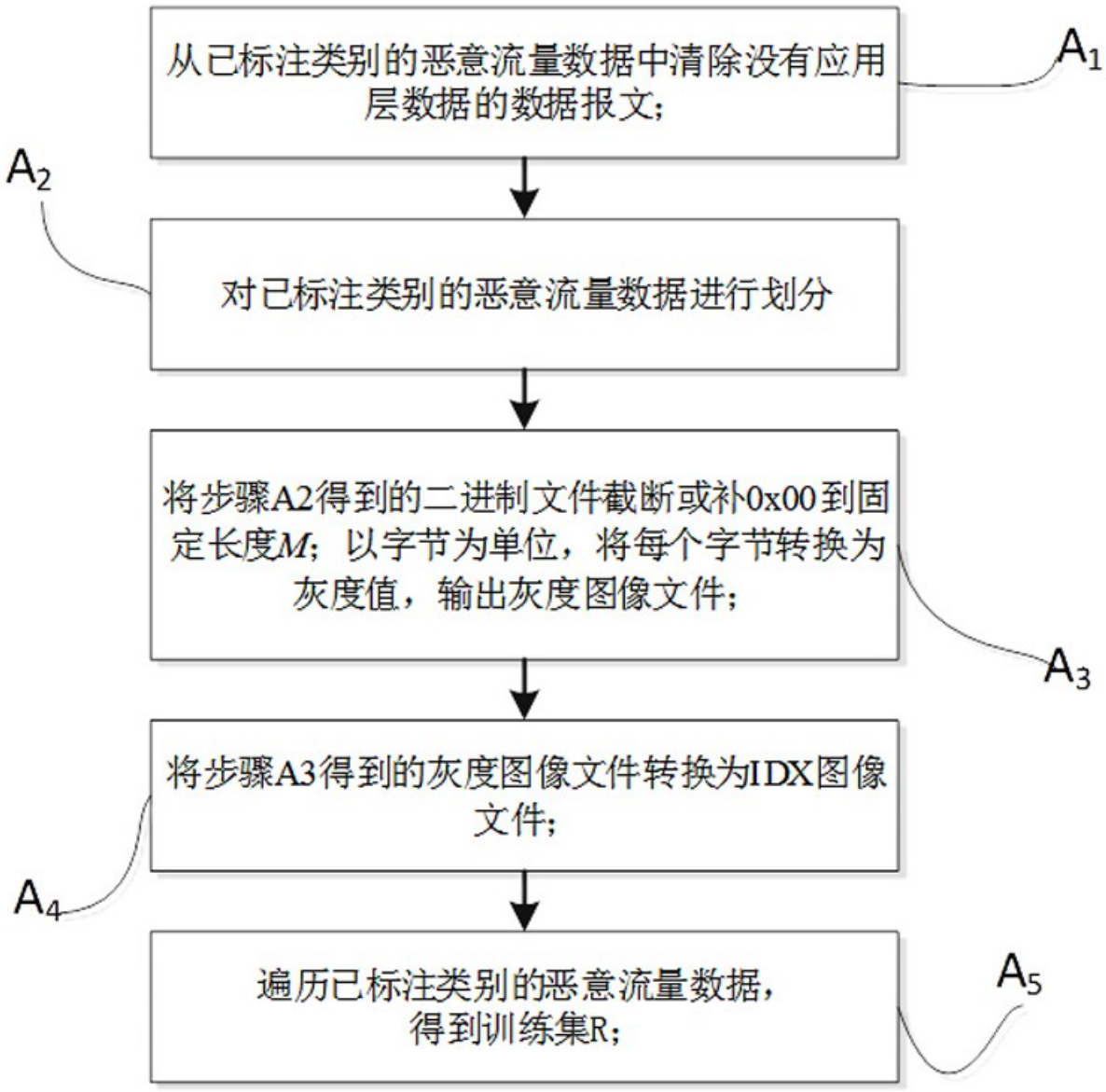

从已标注类别的恶意流量数据中清除没有应用层数据的数据报文；
A1
A2
对已标注类别的恶意流量数据进行划分
将步骤A2得到的二进制文件截断或补0x00到固定长度M；以字节为单位，将每个字节转换为灰度值，输出灰度图像文件；
A3
将步骤A3得到的灰度图像文件转换为IDX图像文件；
A4
遍历已标注类别的恶意流量数据，得到训练集R；
A5

图2

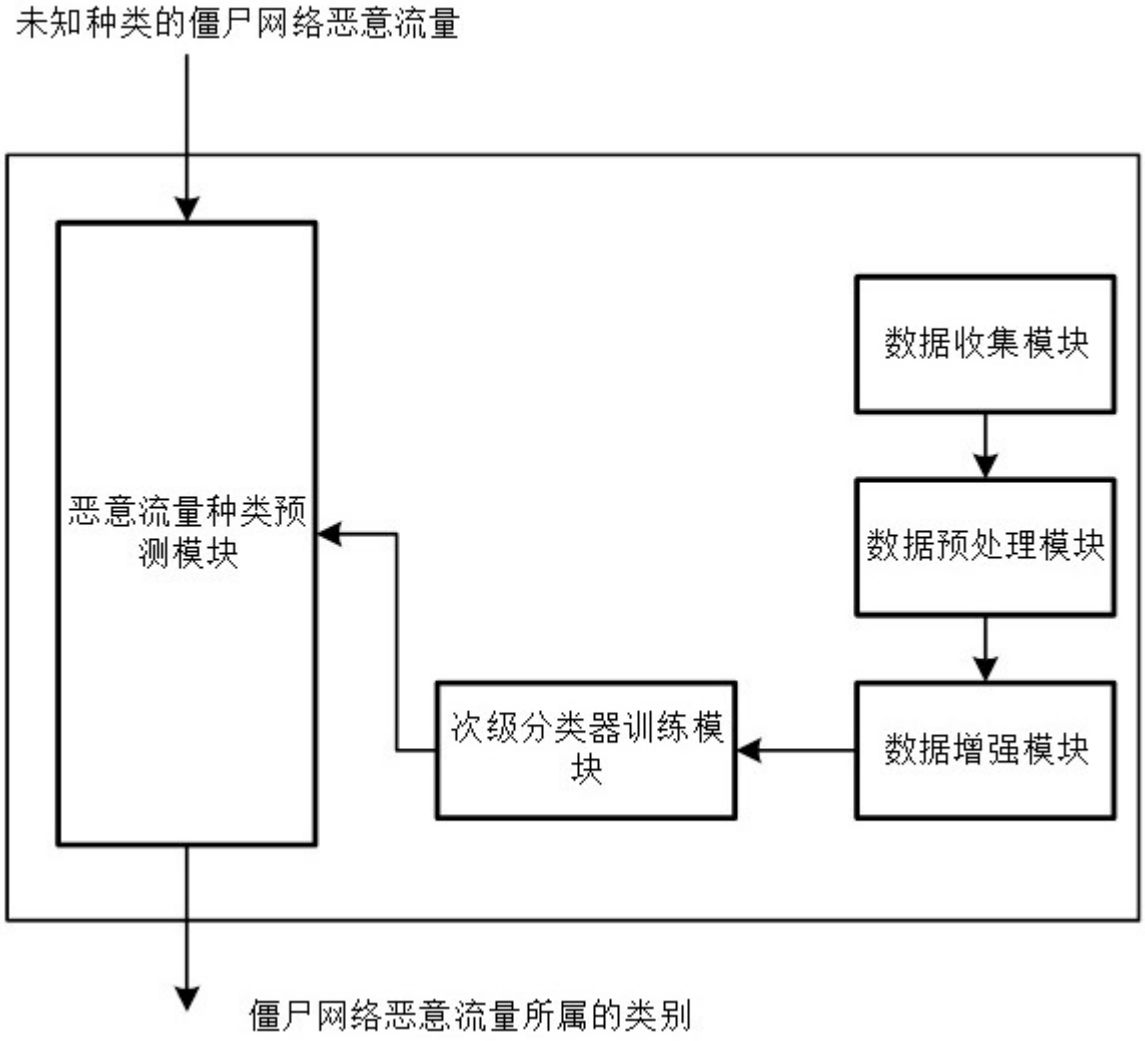

未知种类的僵尸网络恶意流量
数据收集模块
数据预处理模块
数据增强模块
次级分类器训练模块
恶意流量种类预测模块
僵尸网络恶意流量所属的类别

图3