(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 111970309 B
(45)授权公告日 2021.02.02

(21)申请号 202011122116.4
(22)申请日 2020.10.20
(65)同一申请的已公布的文献号
申请公布号 CN 111970309 A
(43)申请公布日 2020.11.20
(73)专利权人 南京理工大学
地址 210094 江苏省南京市玄武区孝陵卫街道孝陵卫街200号
(72)发明人 戚湧 俞建业
(74)专利代理机构 常州佰业腾飞专利代理事务所(普通合伙) 32231
代理人 滕诣迪
(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

(56)对比文件
CN 111428789 A,2020.07.17
CN 110293818 A,2019.10.01
CN 111783442 A,2020.10.16
CN 110881037 A,2020.03.13
US 20200322362 A1,2020.10.08
US 2018288086 A1,2018.10.04
王毅等.基于CNN和LSTM深度网络的伪装用户入侵检测.《计算机科学与探索》.2017,

审查员 徐方南

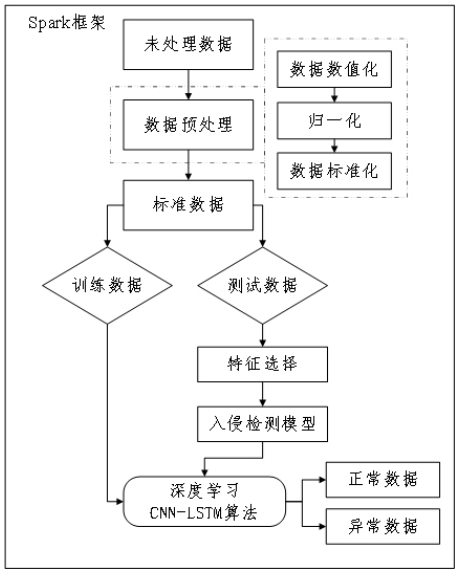
权利要求书3页 说明书7页 附图4页

(54)发明名称

基于Spark车联网组合深度学习入侵检测方法及系统

(57)摘要

本发明公开了一种基于Spark车联网组合深度学习入侵检测方法及系统,包括以下步骤:S1:搭建Spark分布式集群;S2:初始化Spark布式集群并构建CNN-LSTM组合深度学习算法模型以及参数的初始化,将采集到的数据上传到HDFS中;S3:从HDFS上读取数据进行处理并输入进CNN-LSTM组合深度学习算法模型中,对数据进行识别;S4:将数据分成多个RDD进行批量训练,并进行迭代达到预设次数。本发明方法为车联网入侵检测提供了快速准确的检测方法,确保车联网在计算能力受限、应用环境复杂以及存在大量节点网络的情况下能够在短时间内准确、快速的完成入侵检测任务,提供安全可靠通信环境。



CN 111970309 B

1.一种基于Spark车联网组合深度学习入侵检测方法,其特征在于,包括以下步骤:

S1:搭建Spark分布式集群;

Spark分布式集群环境包括一台主节点和多台从节点,包括弹性分布式数据库RDD、分布式文件系统HDFS及容错机制;

S2:初始化Spark布式集群并构建CNN-LSTM组合深度学习算法模型以及参数的初始化,将采集到的数据上传到HDFS中;

所述的步骤S2构建CNN-LSTM组合深度学习算法模型具体为:

构建CNN-LSTM组合深度学习算法模型以及参数的初始化,将采集到的数据上传到HDFS中具体为:

S2.1.1:数据流量采集是对车联网在通信过程中产生的数据交互,包括采集由车载终端从云服务平台获取包括娱乐信息服务、地图、路况、辅助驾驶信息;

S2.1.2:采集由车载终端获取的路测设备的红绿灯信息以及路况和盲区信息;

S2.1.3:采集由车载终端与车载终端传递的信息,包括路况预警信息;

S2.1.4:以及路测设备将路测传感数据上传到云服务平台;其中包括采集协议类型、网络连接状态、网络服务类型;

S2.1.5:采集在数据传输过程中的异常入侵数据;

S2.1.6:在传输节点上连接入侵检测设备对数据进行采集、去重、分析;

将采集到的数据上传到HDFS中,对未处理通信数据进行预处理,采用数据数值化、归一化以及标准化:

S2.2.1:针对未处理的数据进行预处理,具体来说就是发送和接收车辆消息的车载单元OBU以及路侧设施RSU在进行交互的过程中对其中传输的数据进行检测和处理;清洗错误数据和丢失不全的数据,对没有数值化的数据进行数值化,成为有价值的新数据;

S2.2.2:数据数值化,对数据传输过程中各种形式的数据进行数值化操作;就是将S2.2.1中收集到的数据,由原来的字符型转化为数值型数据,更好的分析和识别数据的内容;将属性特征三种协议类型:TCP、UDP、ICMP编码为1,2,3;

S2.2.3:根据数值化的数据,将车联网数据其取值范围在[0,58329]变为(0,1)的小数便于数据更加快速的提取;消除数值化带来的不同量纲的影响;采用0均值标准化,通过数据的均值和标准差进行数据的标准化,经过处理的车联网数据符合标准正态分布,即均值为0,标准差为1,函数原型为:

$$X^* = \frac{X-\mu}{\sigma} \quad (1)$$

其中,μ 为当前车联网采集到的数据的均值,σ 为当前车联网采集到的数据的标准差;

S2.2.4:经过数据的采集以及数据的预处理过程进行车联网数据分析的标准化;

S2.3.1:对数据进行训练时,设置多次的迭代;

S2.3.2:对车联网数据的特征提取由卷积神经网络CNN完成,CNN结构的第一层为网络的输入层;

S2.3.3:第二层为卷积层用于提取特征,使用sigmoid或tanh函数,加快收敛和训练速度;

每个卷积层包含卷积操作和非线性激活两个过程;当前层的特征图可由卷积核进而前一层的输出特征图或原始特征图进行卷积操作得到:

$$X_j^l = \sum_i X_i^{l-1} \otimes K_{ij}^{l-1} + b_j^l \quad (2)$$

其中,X_j^l 表示经过卷积之后第 l 层特征图中的第 j 个位置的输入;X_i^{l-1} 表示第 $l-1$ 层中的第 i 个输入矩阵;K_{ij}^{l-1} 表示在第 l 层和第 $l-1$ 层之间连接第 i 个输入矩阵和第 j 个位置的卷积核;b_j^l 表示到第 l 层特征图和第 j 个位置的偏置量;

S2.3.4:池化层对特征进行抽样,使用LeakyRelu激活:

$$f(x) = \max(0,x) \quad (3)$$

其中,当车联网标准化后的值 x 的取值小于0时,$f(x)$ 的取值为0;当车联网标准化后的值 x 的取值大于0时,$f(x)$ 的取值为 x ;

S2.3.5:最后全连接层将提取的特征连接起来形成整体特征,输入到LSTM中,全连接层的神经元输出计算:

$$y_j^l = \sum_i w_{ij}^l * x_i^{l-1} + b_j^l \quad (4)$$

其中,y_j^l 表示第 l 层全连接层神经元中第 j 个神经元经计算后的输入结果;w_{ij}^l 表示 $l-1$ 层中特征图的第 i 个特征与 l 层中的第 j 个神经元的连接权重;x_i^{l-1} 为 $l-1$ 层中特征图的第 i 个特征值;b_j^l 为第 l 层全连接层神经元中第 j 个神经元的偏置值;

S2.3.6:长短期记忆网络LSTM使用一组gate函数控制反馈,短时间的错误被删除,持续的特性会被保留;使用子网中 $p(t),g(t),f(t)$ 和 $q(t)$ 进行输出,其中 $p(t)$ 为输入门,输入车联网数据特征值的数目;$g(t)$ 为更新门,是指所循环的神经网络的层数,默认为1;$f(t)$ 为遗忘门,是通过权重和偏置以及输入门计算得到的维度;$q(t)$ 为输出门,由权重和偏置以及遗忘门和更新门水平拼接而成的向量计算得到的结果;最后通过两种类型的控制门 σ 和 $\tanh$ 来确定先前学习的反馈 $s(t)$ 和电流输出 $h(t)$:

$$s(t) = \sigma(f(t)) * s(t-1) + \sigma(p(t)) * tanhg(t) \quad (5)$$

$$h(t) = tanhs(t) * \sigma(q(t)) \quad (6)$$

LSTM通过调整网络中的权值和 σ 值来学习输入,从而在输出中有效地生成输入数据之间的时间特征;

S3:从HDFS上读取数据进行处理并输入进CNN-LSTM组合深度学习算法模型中,对数据进行识别;

S4:将数据分成多个RDD进行批量训练,并进行迭代达到预设次数。

2.根据权利要求1所述的基于Spark车联网组合深度学习入侵检测方法,其特征在于,所述的步骤S2.1.4:采集协议类型包括TCP、UDP、ICMP;网络连接状态包括OTH、REJ、RSTO;网络服务类型包括auth、bgp、http、ftp、telent。

3.根据权利要求1所述的基于Spark车联网组合深度学习入侵检测方法,其特征在于,所述的步骤S2.1.5:异常入侵数据包括DoS、Probing、R2L、U2R,具体分类标识包括back、land、neptune、pod、ipsweep、nmap。

4.一种基于Spark车联网组合深度学习入侵检测系统,包括存储器和处理器,存储器存储有计算机程序,其特征在于;所述处理器执行所述计算机程序时实现如权利要求1-3任一所述的方法步骤。

5.一种计算机可读存储介质,其上存储有计算机程序,其特征在于:所述的计算机程序被处理器执行时实现如权利要求1-3任一所述的方法步骤。

基于Spark车联网组合深度学习入侵检测方法及系统

技术领域

[0001] 本发明涉及车联网入侵检测技术领域，具体为一种基于Spark车联网组合深度学习入侵检测方法及系统。

背景技术

[0002] 近年来，随着新兴技术在车联网领域的实践应用，车联网得到更加快速的发展，车与车、与路、与人、与云的通信更加紧密，车联网通信安全对于促进智能交通和智慧城市的发展具有决定性的作用。随着通信能力的提升，大量的网络通信流量也随之而来，但因车联网中计算能力受限、应用环境复杂、分布式多节点和传感网络导致车联网安全问题十分突出，如何确保车联网的安全性，加快车联网落地应用成为汽车厂商和科研人员广泛讨论的话题。因此，利用入侵检测(Intrusion Detection, ID)技术确保车联网通信安全以及识别各种恶意攻击行为成为保障车联网安全的一种重要手段。

[0003] 针对入侵检测的问题，国内外学者提出了多种有效方法，包括机器学习SVM算法、DNN深度神经网络模型、MLP算法模型等，这些算法被用于解决传统入侵检测问题。如Anish Halima等人将SVM方法应用到入侵检测系统(IDS)中。采用SVM和Naive Bayes机器学习算法，应用归一化和特征简约进行分析对比。但是基于机器学习的入侵检测机制的重要缺点是需要大量的训练时间来处理网络先前数据流的大型数据集，在处理大数据网络环境中，尤其是复杂的车联网当中对检测时间极为重要。R. Vinayakumar等人提出混合深度神经网络(DNN)模型用于检测和分类未知的网络攻击。丁红卫等人提出基于深度卷积神经网络的入侵检测方法，将网络数据转换为图像并进行降维。通过训练和识别从而提高检测的准确率、误报率和检测速率。

[0004] 但是以上这些算法不能直接应用在车联网的实际环境中，一是车联网结构复杂，不仅是车与自身需要通信，车与人、与车、与路以及与云都需要交互；二是网络通信协议和方式众多，不仅有蓝牙、无线、有线还有移动蜂窝网络和LTE-V2X；三是网络拓扑变化快，车辆是处于高速移动过程当中的，所以车联网的网络拓扑也是根据实际的环境在不断变化的。

[0005] 针对上述问题分析以及车联网实际特点，为了解决车联网通信过程当中网络流量巨大难以快速有效检测，以及入侵检测准确率的问题，需要提出一种新的应用于车辆网当中的入侵检测解决方案。

发明内容

[0006] 1、本发明的目的

[0007] 本发明提供一种基于Spark的分布式组合深度学习算法的车联网入侵检测方法及系统，以提高车联网入侵检测的时间和检测率，解决在车联网应用当中通信网络安全问题。

[0008] 2、本发明所采用的技术方案

[0009] 本发明公开了一种基于Spark车联网组合深度学习入侵检测方法，包括以下步骤：

[0010] S1:搭建Spark分布式集群;

[0011] S2:初始化Spark布式集群并构建CNN-LSTM组合深度学习算法模型以及参数的初始化,将采集到的数据上传到HDFS中;

[0012] 所述的步骤S2构建CNN-LSTM组合深度学习算法模型具体为:

[0013] S2.3.1:对数据进行训练时,设置100次的迭代;

[0014] S2.3.2:对车联网数据的特征提取由卷积神经网络CNN完成,CNN结构的第一层为网络的输入层,本层的网络的输入维度为11*11;

[0015] S2.3.3:第二层为卷积层用于提取特征,使用sigmoid或tanh函数,加快收敛和训练速度,本层的卷积核大小为[5*5];

[0016] 每个卷积层包含卷积操作和非线性激活两个过程;当前层的特征图可由卷积核进而前一层的输出特征图或原始特征图进行卷积操作得到:

[0017] $$X_j^l = \sum_i X_i^{l-1} \otimes K_{ij}^{l-1} + b_j^l \quad (2)$$

[0018] 其中,X_j^l表示经过卷积之后第l层特征图中的第j个位置的输入;X_i^{l-1}表示第$l-1$层中的第i个输入矩阵;K_{ij}^{l-1}表示在第l层和第$l-1$层之间连接第i个输入矩阵和第j个位置的卷积核;b_j^l表示到第l层特征图和第j个位置的偏置量;

[0019] S2.3.4:池化层对特征进行抽样,使用LeakyRelu激活:

[0020] $$f(x) = \max(0,x) \quad (3)$$

[0021] 其中,当车联网标准化后的值x的取值小于0时,$f(x)$的取值为0;当车联网标准化后的值x的取值大于0时,$f(x)$的取值为x;

[0022] S2.3.5:最后全连接层将提取的特征连接起来形成整体特征,输入到LSTM中,全连接层的神经元输出计算:

[0023] $$y_j^l = \sum_i w_{ij}^l * x_i^{l-1} + b_j^l \quad (4)$$

[0024] 其中,y_j^l表示第l层全连接层神经元中第j个神经元经计算后的输入结果;w_{ij}^l表示$l-1$层中特征图的第i个特征与l层中的第j个神经元的连接权重;x_i^l为$l-1$层中特征图的第i个特征值;b_j^l为第l层全连接层神经元中第j个神经元的偏置值;

[0025] S2.3.6:长短期记忆网络LSTM使用一组gate函数控制反馈,短时间的错误被删除,持续的特性会被保留;使用子网中$p(t),g(t),f(t)$和$q(t)$进行输出,其中$p(t)$为输入门,输入车联网数据特征值的数目;$g(t)$为更新门,是指所循环的神经网络的层数,默认为1;

$f(t)$为遗忘门，是通过权重和偏置以及输入门计算得到的维度；$q(t)$为输出门，由权重和偏置以及遗忘门和更新门水平拼接而成的向量计算得到的结果；最后通过两种类型的控制门σ和$\tanh$来确定先前学习的反馈$s(t)$和电流输出$h(t)$：

[0026] $$s(t) = \sigma(f(t)) * s(t-1) + \sigma(p(t)) * tanhg(t) \quad (5)$$

[0027] $$h_t = tanhs(t) * \sigma(q(t)) \quad (6)$$

[0028] LSTM通过调整网络中的权值和σ值来学习输入，从而在输出中有效地生成输入数据之间的时间特征；

[0029] S3：从HDFS上读取数据进行处理并输入进CNN-LSTM组合深度学习算法模型中，对数据进行识别；

[0030] S4：将数据分成多个RDD进行批量训练，并进行迭代达到预设次数。

[0031] 所述的步骤S1：Spark分布式集群环境包括一台主节点和多台从节点，包括弹性分布式数据库RDD、分布式文件系统及容错机制。

[0032] 所述的步骤S2，构建CNN-LSTM组合深度学习算法模型以及参数的初始化，将采集到的数据上传到HDFS中具体为：

[0033] S2.1.1：数据流量采集主要是对车联网在通信过程中产生的数据交互，包括正常交互数据，由车载终端从云服务平台获取包括娱乐信息服务、地图、路况、辅助驾驶；

[0034] S2.1.2：由车载终端获取路测设备的红绿灯信息以及路况和盲区信息；

[0035] S2.1.3：由车载终端与车载终端进行信息传递包括路况预警信息；

[0036] S2.1.4：以及路测设备将路测传感数据或计算上传到云服务平台；包括采集协议类型、网络连接状态、网络服务类型；

[0037] S2.1.5：在数据传输过程中的异常入侵数据；

[0038] S2.1.6：在传输节点上连接入侵检测设备对数据进行采集、去重、分析。

[0039] 所述的步骤S2.1.4：采集协议类型包括TCP、UDP、ICMP；网络连接状态包括OTH、REJ、RST0；网络服务类型包括auth、bgp、http、ftp、telent。

[0040] 所述的步骤S2.1.5：在数据传输过程中的异常入侵数据，包括DoS、Probing、R2L、U2R，具体分类标识包括back、land、neptune、pod、ipsweep、nmap。

[0041] 所述的步骤S2，将采集到的数据上传到HDFS中，对未处理通信数据进行预处理，采用数据数值化、归一化以及标准化：

[0042] S2.2.1：针对未处理的数据进行预处理，具体来说就是发送和接收车辆消息的车载单元OBU以及路侧设施RSU在进行交互的过程中对其中传输的数据进行检测和处理；清洗错误数据和丢失不全的数据，对没有数值化的数据进行数值化，成为有价值的新数据；

[0043] S2.2.2：数据数值化，对数据传输过程中各种形式的数据进行数值化操作；就是将S2.2.1中收集到的数据，由原来的字符型转化为数值型数据，更好的分析和识别数据的内容；例如将属性特征三种协议类型：TCP、UDP、ICMP编码为1，2，3；

[0044] S2.2.3：根据数值化的数据，将车联网数据其取值范围在[0,58329]变为(0,1)或者(1,1)之间的小数便于数据更加快速的提取；消除数值化带来的不同量纲的影响；采用0均值标准化，通过数据的均值和标准差进行数据的标准化，经过处理的车联网数据符合标

准正态分布，即均值为0，标准差为1，函数原型为：

[0045] $$X^* = \frac{X-\mu}{\sigma} \quad (1)$$

[0046] 其中，μ 为当前车联网采集到的数据的均值，σ 为当前车联网采集到的数据的标准差；

[0047] S2.2.4:经过数据的采集以及数据的预处理过程进行车联网数据分析的标准化。

[0048] 本发明公开了一种基于Spark车联网组合深度学习入侵检测系统，包括存储器和处理器，存储器存储有计算机程序，所述处理器执行所述计算机程序时实现所述的方法步骤。

[0049] 本发明公开了一种计算机可读存储介质，其上存储有计算机程序，所述的计算机程序被处理器执行时实现所述的方法步骤。

[0050] 3、本发明所采用的有益效果

[0051] (1)本发明使用组合深度学习算法用于车联网入侵检测的检测结果，提高车联网入侵检测的检测准确率，使用的组合算法计算简单，易于实现，具有实际应用价值；

[0052] (2)本发明通过搭建Spark分布式集群，减少入侵检测所需实际，具有良好的实时性；

[0053] (3)本发明能精准捕捉车联网数据通信过程中的异常数据，在最快的时间内完成检测，得到检测结果，具有较快的速度；

[0054] (4)本发明在车联网体系结构中的各个部分都可以使用，使用的检测算法也可以进行替换，具有良好的可迁移性和可扩展性。

附图说明

[0055] 图1为本发明基于Spark平台车联网分布式组合深度学习算法流程图。

[0056] 图2为组合深度学习算法模型示意图。

[0057] 图3为本发明和部分已有方法在准确率和误报率比较图。

[0058] 图4为本发明和部分已有方法检测时间比较图。

[0059] 图5为本发明的整体流程图。

具体实施方式

[0060] 下面结合本发明实例中的附图，对本发明实例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明的实施例，本领域技术人员在没有做创造性劳动前提下所获得的所有其他实施例，都属于本发明的保护范围。

[0061] 下面将结合附图对本发明实例作进一步地详细描述。

[0062] 如图1、5所示，基于Spark分布式集群和组合深度学习算法模型对车联网数据流量进行入侵检测，得到结果。

[0063] S1:通过搭建Spark分布式集群将所述步骤3的组合深度学习方法用于车联网入侵检测。Spark分布式集群环境包括一台主节点和四台从节点，Spark架构主要用到弹性分布

式数据库RDD、分布式文件系统及容错机制等模块；

[0064] S2：初始化Spark并构建所述的CNN-LSTM组合深度学习算法模型以及参数的初始化，将采集到的数据上传到HDFS中；

[0065] S3：从HDFS上读取数据进行处理并输入进CNN-LSTM组合深度学习算法模型中，对数据进行识别，见步骤3。

[0066] S4：将数据分成多个RDD进行批量训练，并进行迭代达到预设次数。

[0067] 所述的S2中，对车联网通信过程的原始数据进行采集；

[0068] S2.1.1：数据流量采集主要是对车联网在通信过程中产生的数据交互，包括正常交互数据，由车载终端从云服务平台获取的娱乐信息服务、地图、路况、辅助驾驶等；

[0069] S2.1.2：由车载终端获取路测设备的红绿灯信息以及路况和盲区信息；

[0070] S2.1.3：由车载终端与车载终端进行信息传递和路况预警等信息；

[0071] S2.1.4：以及路测设备将路测传感数据或者高复杂度的计算上传到云服务平台当中等。共包括采集3种协议类型(TCP、UDP、ICMP)；11种网络连接状态，如：OTH、REJ、RSTO等；70种网络服务类型，包括auth、bgp、http、ftp、telent等

[0072] S2.1.5：另外，在数据传输过程中的异常入侵数据，包括DoS(拒绝服务)攻击、Probing(探测攻击)、R2L(远程非法访问)、U2R(越权访问)等，具体分类标识为back、land、neptune、pod、ipsweep、nmap等共计4大类39种攻击类型。

[0073] S2..6：对车联网通信过程中的数据流量进行采集操作如下：所有交互数据不论采用怎样的交互手段最终都会通过光纤等有线的方式进行汇聚，因此在传输节点(交换机或路由器)上连接入侵检测设备对数据进行采集、去重、分析。

[0074] 步骤2.2：对未处理通信数据进行预处理，采用数据数值化、归一化以及标准化的步骤进行；

[0075] S2.2.1：针对未处理的数据进行预处理，具体来说就是发送和接收车辆消息的车载单元(OBU)以及路侧设施(RSU)在进行交互的过程中对其中传输的数据进行检测和处理。清洗错误数据和丢失不全的数据，对没有数值化的数据进行数值化，成为有价值的新数据。

[0076] S2.2.2：数据数值化，对数据传输过程中各种形式的数据进行数值化操作。就是将步骤1中收集到的数据，由原来的字符型转化为数值型数据，更好的分析和识别数据的内容。例如将属性特征三种协议类型：TCP、UDP、ICMP编码为1，2，3。

[0077] S2.2.3：根据数值化的数据，将车联网数据其取值范围在[0,58329]变为(0,1)或者(1,1)之间的小数便于数据更加快速的提取。消除数值化带来的不同量纲的影响。采用0均值标准化，通过数据的均值和标准差进行数据的标准化，经过处理的车联网数据符合标准正态分布，即均值为0，标准差为1，函数原型为：

[0078] $$X^* = \frac{X-\mu}{\sigma} \quad (1)$$

[0079] 其中，μ 为当前车联网采集到的数据的均值，σ 为当前车联网采集到的数据的标准差。

[0080] S2.2.4：经过数据的采集以及数据的预处理过程进行车联网数据分析的标准化。

[0081] S2.3：基于CNN-LSTM组合深度学习算法模型进行数据训练和特征提取检测数据流

量；

[0082] S2.3.1:对数据进行训练时，设置100次的迭代；

[0083] S2.3.2:对车联网数据的特征提取由卷积神经网络(CNN)完成，CNN结构的第一层为网络的输入层，本层的网络的输入维度为11*11；

[0084] S2.3.3:第二层为卷积层用于提取特征，使用sigmoid或tanh函数，加快收敛和训练速度，本层的卷积核大小为[5*5]；

[0085] 每个卷积层包含卷积操作和非线性激活两个过程。当前层的特征图可由卷积核进而前一层的输出特征图或原始特征图进行卷积操作得到：

[0086] $$X_j^l = \sum_i X_i^{l-1} \otimes K_{ij}^{l-1} + b_j^l \quad (2)$$

[0087] 其中，X_j^l表示经过卷积之后第l层特征图中的第j个位置的输入；X_i^{l-1}表示第$l-1$层中的第i个输入矩阵；K_{ij}^{l-1}表示在第l层和第$l-1$层之间连接第i个输入矩阵和第j个位置的卷积核；b_j^l表示到第l层特征图和第j个位置的偏置量；

[0088] S2.3.4:池化层对特征进行抽样，使用LeakyRelu激活：

[0089] $$f(x) = \max(0,x) \quad (3)$$

[0090] 其中，当车联网标准化后的值x的取值小于0时，$f(x)$的取值为0；当车联网标准化后的值x的取值大于0时，$f(x)$的取值为x；

[0091] S2.3.5:最后全连接层将提取的特征连接起来形成整体特征，输入到LSTM中，全连接层的神经元输出计算：

[0092] $$y_j^l = \sum_i w_{ij}^l * x_i^{l-1} + b_j^l \quad (4)$$

[0093] 其中，y_j^l表示第l层全连接层神经元中第j个神经元经计算后的输入结果；w_{ij}^l表示$l-1$层中特征图的第i个特征与l层中的第j个神经元的连接权重；x_i^l为$l-1$层中特征图的第i个特征值；b_j^l为第l层全连接层神经元中第j个神经元的偏置值。

[0094] S2.3.6:长短期记忆网络(LSTM)使用一组gate函数控制反馈，短时间的错误被删除，持续的特性会被保留。使用子网中($p(t),g(t),f(t)$和$q(t)$)进行输出，其中$p(t)$是指输入门，是输入的车联网数据特征值的数目；$g(t)$是指更新门，是指所循环的神经网

络的层数，这里默认是1；$f(t)$是指遗忘门，具体是通过权重和偏置以及输入门计算得到的维度；$q(t)$是指输出门，具体是由权重和偏置以及遗忘门和更新门水平拼接而成的向量计算得到的结果。最后通过两种类型的控制门（σ和$\tan h$）来确定先前学习的反馈$s(t)$和电流输出$h(t)$：

[0095] $$s(t) = \sigma(f(t)) * s(t-1) + \sigma(p(t)) * tanhg(t) \quad (5)$$

[0096] $$h_t = tanhs(t) * \sigma(q(t)) \quad (6)$$

[0097] LSTM通过调整网络中的权值和σ值来学习输入，从而在输出中有效地生成输入数据之间的时间特征。

[0098] 3、为了验证本发明方法相较于现有技术具有较好的效果，使用相关数据集对方法进行对比验证。如图3为本发明和部分已有方法在NSL-KDD数据集上的比较结果。其中上图总共使用148517条数据，下图在UNSW-NB15数据集上比较，总共使用121981条数据，通过对四种不同方法与本方法的比较，组合深度学习算法模型分别达到了99.7%和99.4%的准确率。

[0099] 4、为了验证本发明方法相较于现有技术具有更低的检测时间，使用Spark分布式集群和相关数据集对组合深度学习算法模型进行对比验证。如图4为本发明和部分已有方法在NSL-KDD和UNSW_NB15数据集上的比较结果。可以看出基于Spark的分布式组合深度学习方法具有最低的检测时间。

[0100] 以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明披露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求书的保护范围为准。

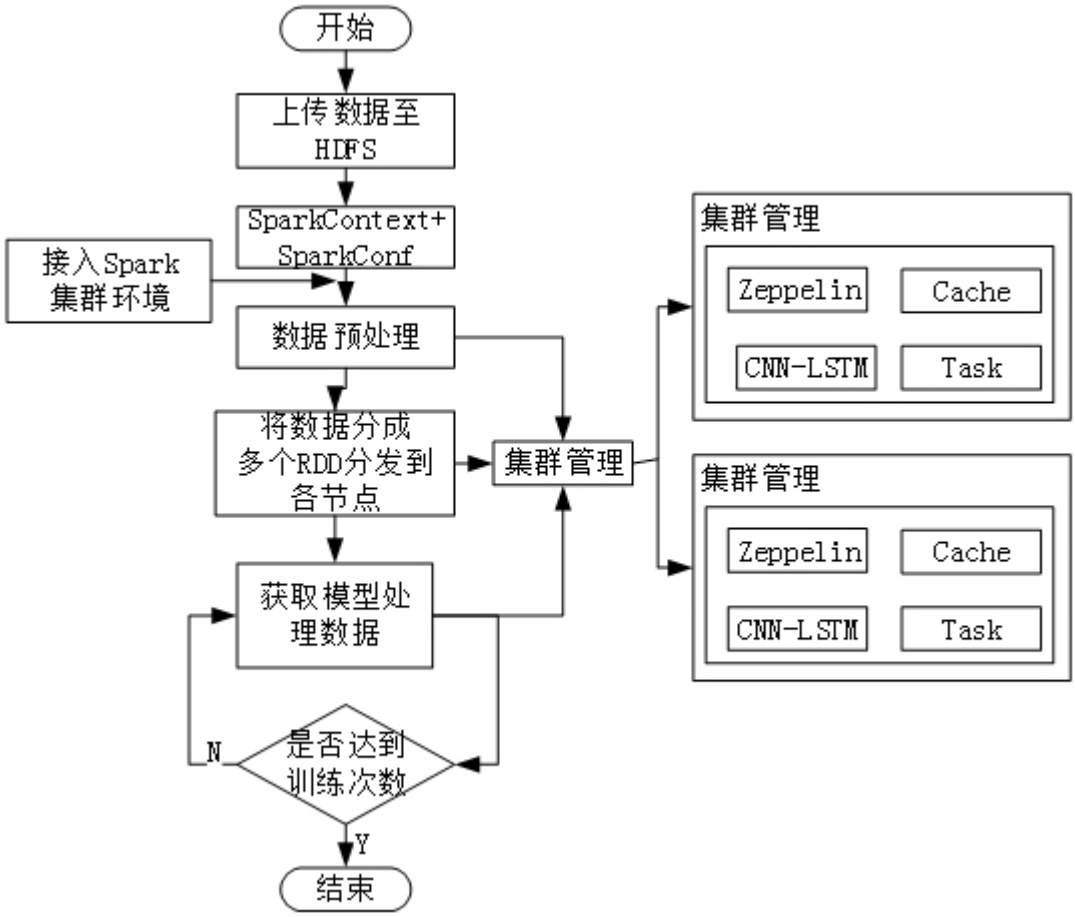

开始
上传数据至HDFS
SparkContext+SparkConf
接入Spark集群环境
数据预处理
将数据分成多个RDD分发到各节点
获取模型处理数据
是否达到训练次数
N
Y
结束
集群管理
集群管理
集群管理
Zeppelin
Cache
CNN-LSTM
Task
Zeppelin
Cache
CNN-LSTM
Task

图1

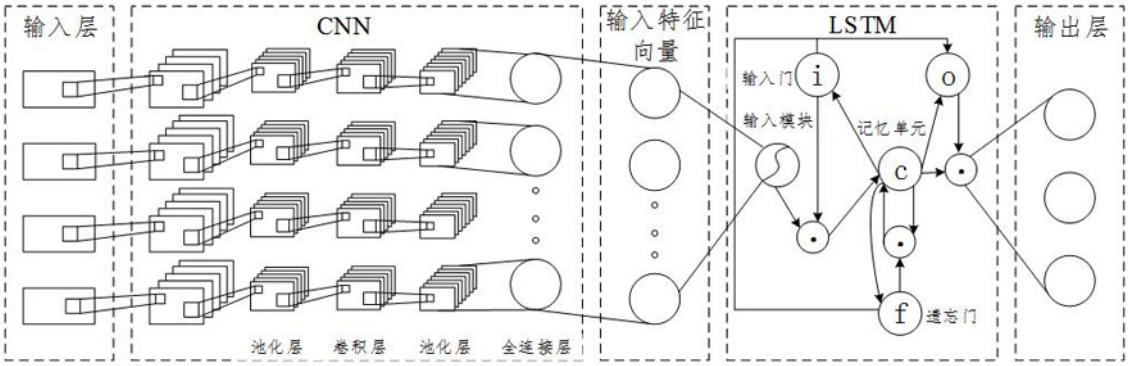

输入层
CNN
池化层
卷积层
池化层
全连接层
输入特征向量
LSTM
输入门
输入模块
记忆单元
遗忘门
输出层

图2

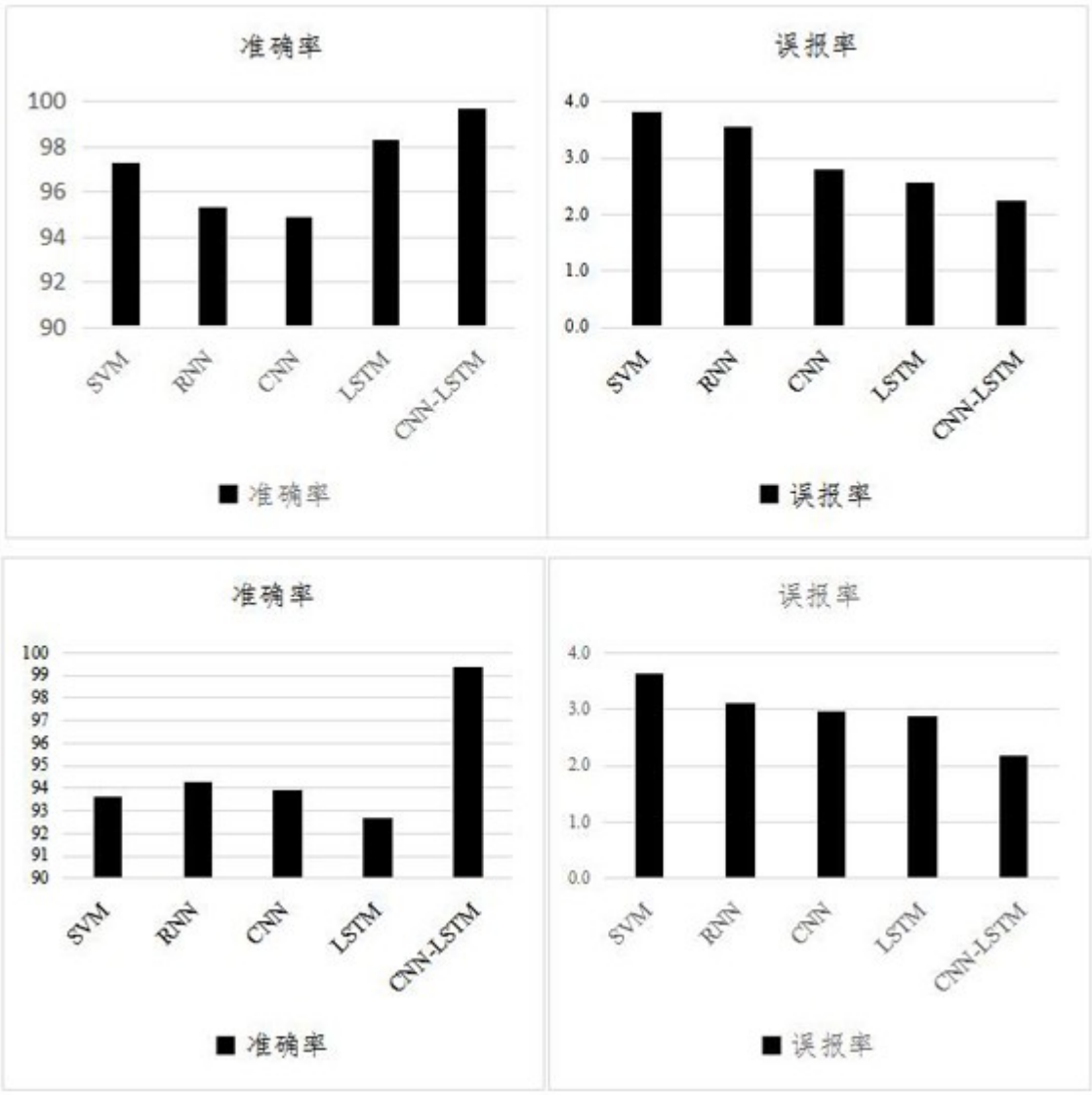

准确率
100
98
96
94
92
90
SVM
RNN
CNN
LSTM
CNN-LSTM
准确率
误报率
4.0
3.0
2.0
1.0
0.0
SVM
RNN
CNN
LSTM
CNN-LSTM
误报率
准确率
100
99
98
97
96
95
94
93
92
91
90
SVM
RNN
CNN
LSTM
CNN-LSTM
准确率
误报率
4.0
3.0
2.0
1.0
0.0
SVM
RNN
CNN
LSTM
CNN-LSTM
误报率

图3

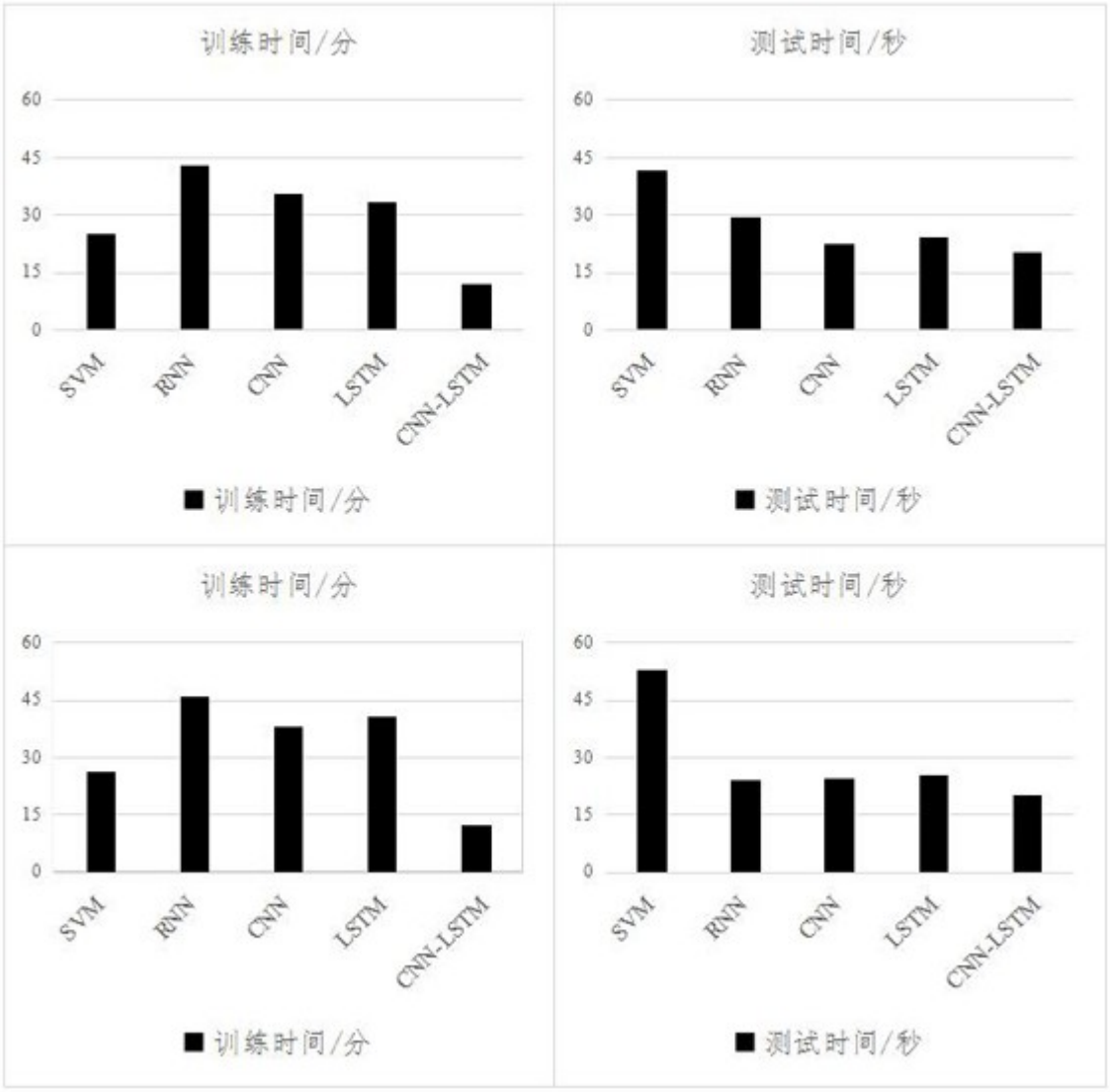

训练时间/分
60
45
30
15
0
SVM
RNN
CNN
LSTM
CNN-LSTM
训练时间/分
测试时间/秒
60
45
30
15
0
SVM
RNN
CNN
LSTM
CNN-LSTM
测试时间/秒
训练时间/分
60
45
30
15
0
SVM
RNN
CNN
LSTM
CNN-LSTM
训练时间/分
测试时间/秒
60
45
30
15
0
SVM
RNN
CNN
LSTM
CNN-LSTM
测试时间/秒

图4

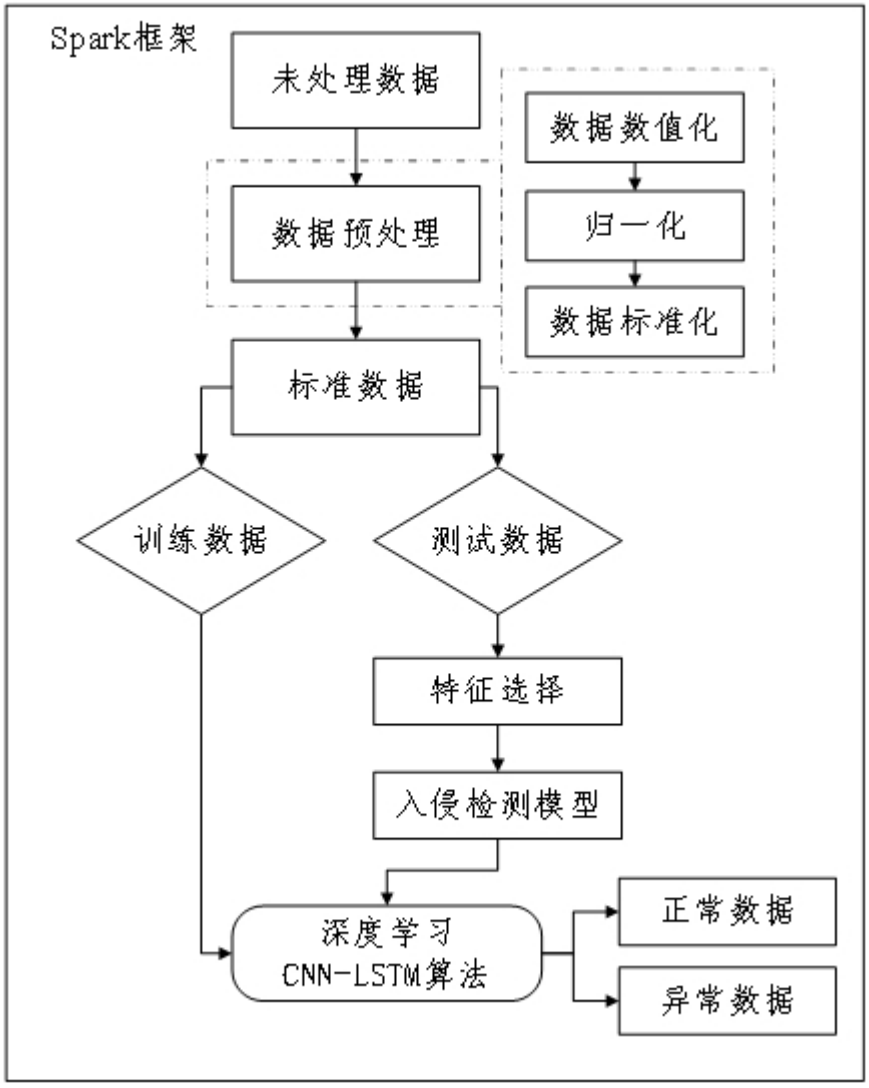

Spark框架
未处理数据
数据预处理
标准数据
数据数值化
归一化
数据标准化
训练数据
测试数据
特征选择
入侵检测模型
深度学习
CNN-LSTM算法
正常数据
异常数据

图5