(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113344562 A
(43)申请公布日 2021.09.03

(21)申请号 202110905722.1
(22)申请日 2021.08.09
(71)申请人 四川大学
地址 610065 四川省成都市武侯区一环路南一段24号
(72)发明人 王海舟 文廷科 肖元星 韩莉君 王安琪
(74)专利代理机构 成都禾创知家知识产权代理有限公司 51284
代理人 刘凯
(51)Int.Cl.
G06Q 20/06(2012.01)
G06Q 20/38(2012.01)
G06Q 20/40(2012.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

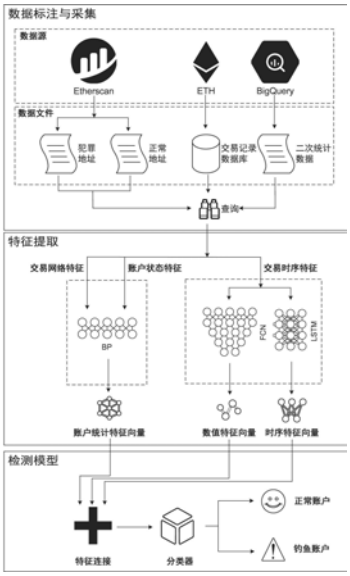
权利要求书3页 说明书16页 附图5页

(54)发明名称

基于深度神经网络的以太坊钓鱼诈骗账户检测方法与装置

(57)摘要

本发明公开了一种基于深度神经网络的以太坊钓鱼诈骗账户检测方法与装置,首先有目标性地采集了权威网站Etherscan上的钓鱼诈骗账户列表用以标注账户类别,然后基于这些钓鱼诈骗账户构建了以太坊钓鱼诈骗子网络,并从中整理得到数据集ETHScam;ETHScam针对以太坊账户以及账户所参与的所有交易记录提取了15个特征,包括账户状态特征、账户交易网络特征和账户交易序列特征三个类别;最终,提出一个以太坊钓鱼诈骗账户检测模型MFL。该模型使用FCN和LSTM来提取账户交易序列特征的数值特征向量和时序特征向量,并结合BP神经网络学习账户状态特征和账户交易网络特征得到账户的统计特征向量,实现了对账户的分类。



CN 113344562 A

1.一种基于深度神经网络的以太坊钓鱼诈骗账户检测方法,其特征在于,包括以下步骤:

步骤1:通过网络爬虫和以太坊节点,获取账户的地址、标记和交易的相关字段,构建以太坊钓鱼诈骗二阶子网络;从中分析并提取出以太坊钓鱼诈骗的账户交易序列特征、账户状态特征和账户交易网络特征,构建以太坊钓鱼诈骗账户数据集ETHScam;

步骤2:构建一个基于FCN-LSTM网络和BP神经网络的深度学习模型,模型命名为MFL,根据输入的以太坊钓鱼诈骗账户数据集ETHScam进行特征提取:将账户交易序列特征投入FCN和LSTM并置的网络中提取出交易的数值特征向量和时序特征向量,将账户状态特征和账户交易网络特征投入BP神经网络学习得到统计特征向量;

步骤3:将统计特征向量、数值特征向量和时序特征向量进行拼接,然后将其输入到全连接神经网络构建的分类器中进行分类,得到账户是否为钓鱼账户的分类结果。

2.根据权利要求1所述的基于深度神经网络的以太坊钓鱼诈骗账户检测方法,其特征在于,所述以太坊钓鱼诈骗二阶子网络具体包括:通过编写爬虫程序从网站Etherscan上获取标注数据;使用Bigquery服务快速查找得到需要的以太坊区块、账户和交易的统计数据;在本地服务器运行一个以太坊节点,实时与以太坊主网络同步,通过查询本地全节点同步的以太坊数据,以Etherscan标记的钓鱼诈骗账户为起点,向外枚举扩展邻点构建以太坊钓鱼诈骗二阶子网络。

3.根据权利要求1所述的基于深度神经网络的以太坊钓鱼诈骗账户检测方法,其特征在于,步骤1中,所述构建以太坊钓鱼诈骗账户数据集ETHScam具体包括:

步骤1.1:提取账户交易序列特征:从太坊钓鱼诈骗二阶子网络中选择已经被标记的账户作为钓鱼诈骗账户,随机选择出未被标记的账户作为正常账户;对于被选择的钓鱼诈骗账户和正常账户,先从子网络数据中取出其参与的所有交易记录,然后提取出每个交易中的交易时间戳、交易以太币数目、交易手续费与转账方向共计4个字段;对于交易手续费*GasPrice*,计算其与块内平均交易手续费*AvgGasPrice*的比值*GasPriceRatio*,计算公式为:

$$GasPriceRatio = \frac{GasPrice}{AvgGasPrice}$$

其中,通过Bigquery的SQL查询功能获取块内平均交易手续费*AvgGasPrice*数据;

步骤1.2:提取账户状态特征:基于Bigquery和上一步骤获取到的交易记录,计算得到账户目前的状态信息,具体为通过Bigquery查询到指定账户目前的余额;计算参与的交易数据得到账户接收和转出的以太币数目以及转出以太币数目与接收以太币数目的比值;

步骤1.3:提取账户交易网络特征:将账户参与的所有交易按照交易方向划分为转入交易和转出交易两类,统计两类交易的数目得到转入账户数目、转出账户数目以及转入转出账户数目比值;再计算两类交易的平均转账以太币数目得到平均转入以太币数目、平均转出以太币数目以及平均转入转出以太币数目比值。

4.根据权利要求1所述的基于深度神经网络的以太坊钓鱼诈骗账户检测方法,其特征在于,所述账户交易序列特征包括交易时间戳、交易以太币数目、交易方向和交易手续费比值;所述账户状态特征包括账户余额、账户涉及交易数量、接收以太币数目、转出以太币数目和以太币转出接收比值;所述账户交易网络特征包括转入账户数目、转出账户数目、转入

转出账户数目比值、平均转入以太币数目、平均转出以太币数目和平均转入转出以太币数目比值。

5.根据权利要求1所述的基于深度神经网络的以太坊钓鱼诈骗账户检测方法，其特征在于，所述深度学习模型包括：

输入层：

所述输入层用于输入预处理之后得到的账户交易时间序列、账户状态特征和账户交易网络特征；

所述输入层分三部分，第一部分将经过预处理后的账户交易时间序列TS作为输入$TS=\{TS_1,TS_2,TS_3,...,TS_n\}$；交易时间序列预处理为对原始账户交易时间序列$TS_0$使用滑动窗口采样法采样并进行归一化得到$TS$；本部分输出将会投入到特征提取层中用于提取账户交易时序序列的时序特征向量和数值特征向量；

输入层的第二部分和第三部分将账户状态特征和账户交易网络特征的统计特征向量直接并置到时序特征向量和数值特征向量之后，作为分类器的输入；

特征提取：

所述特征提取包括两大模块，分别为基于全卷积神经网络FCN的第一特征提取模块和基于LSTM的第二特征提取模块；第一特征提取模块将经过预处理后的账户交易时间序列作为输入，投入到全卷积神经网络中处理后，经过一个全局池化层得到时序变量的内部隐含特征M，M是32维的数值特征向量；第二特征提取模块用于提取时序特征，其包括8个细胞，输入层Dropout率设置为0.2，隐藏层Dropout率设置为0.5；最终输出8维的时序特征向量T；

所述账户状态特征和账户交易网络特征均为针对账户的统计特征，将此两个部分的特征向量分别进行归一化然后输入到BP神经网络中，最终得到16维的统计特征向量S；

特征拼接：

将统计特征向量S、时序特征向量T和数值特征向量M拼接得到账户特征表示向量$V=\{V_1,V_2,V_3,...,V_{56}\}$，其表示为：

$$V=T\oplus S\oplus M$$

输出层：

将拼接起来的统计特征向量、时序特征向量和数值特征向量投入到全连接神经网络，然后通过Sigmoid函数计算得到该账户是钓鱼账户的概率，以此得到最终的分类结果P_d，其表示为：

$$p_d=sigmoid(V_E)$$

其中，V_E为最终判断账户是否为钓鱼账户的向量，经过Sigmoid函数得到预测结果；

模型的优化目标为最小化交叉熵损失函数L，其表示为：

$$L=-\sum_{d\in D}[y_d log(p_d)+(1-y_d)log(1-p_d)]$$

其中，d表示样本，D表示样本数据集；y_d表示样本的真实值，p_d为样本的预测值。

6.根据权利要求5所述的基于深度神经网络的以太坊钓鱼诈骗账户检测方法，其特征在于，所述全卷积神经网络FCN包括三个时间卷积块用作特征提取器；所述卷积块包括具有多个滤波器的卷积层和多个内核，每一个卷积层都经过批量归一化；批量规范化层后接

ReLU激活函数;且前两个卷积块以一个压缩和激励块结束,所有压缩和激励块的衰减率r设置为16;最后一个卷积块后接一个全局平均池化层;所述压缩和激励块带来的附加参数的总数为:

$$P=\frac{2}{r}\sum_{S=1}^{S}R_S\cdot G_S^2$$

其中,P是附加参数的总数,r表示衰减率,S表示阶段数,G_s表示阶段S的输出特征图的数目,R_S表示阶段S的重复块数。

7.一种基于深度神经网络的以太坊钓鱼诈骗账户检测装置,其特征在于,包括数据标注与采集模块、特征提取模块和检测模块;

所述数据标注与采集模块通过网络爬虫和以太坊节点,获取账户的地址、标记和交易的相关字段,构建以太坊钓鱼诈骗子网络;从中分析并提取出以太坊钓鱼诈骗的账户交易序列特征、账户状态特征和账户交易网络特征,构建以太坊钓鱼诈骗账户数据集ETHScam;

所述特征提取模块通过FCN和LSTM并置的网络从账户交易序列特征中提取出交易的数值特征向量和时序特征向量;通过BP神经网络从账户状态特征和账户交易网络特征中提取统计特征向量;

所述检测模块将统计特征向量、数值特征向量和时序特征向量进行拼接,再通过全连接神经网络构建的分类器进行分类。

基于深度神经网络的以太坊钓鱼诈骗账户检测方法与装置

技术领域

[0001] 本发明涉及网络安全技术领域，具体为一种基于深度神经网络的以太坊钓鱼诈骗账户检测方法与装置。

背景技术

[0002] 区块链技术是以比特币、以太坊为代表的众多加密货币方案的底层核心技术，最初设计目的是解决电子支付中过度依赖可信第三方的问题。区块链组合使用P2P网络、分布式计算等成熟技术，并结合哈希函数、非对称密码、数字签名和零知识证明等密码学技术，成为一种全新的分布式基础架构和计算范式。区块链技术极具应用潜力，其应用范围已从最初的加密货币延伸至金融、物联网、智能制造等多个领域，引起了工业界、学术界和国家层面的广泛关注。世界经济论坛对区块链在金融场景下的应用进行了预测分析，认为区块链将在跨境支付、保险、贷款等多方面重塑金融市场基础设施。

[0003] 随着理论研究的深入，区块链在不断持续展现出蓬勃生命力的同时，其自身的安全问题逐渐显露。针对加密货币应用的安全威胁以及针对区块链平台的各种犯罪行为呈现高发态势。在交易平台被盗事件频发、智能合约漏洞凸显、利用匿名交易实施犯罪等威胁之外，借助区块链加密货币实施的钓鱼诈骗犯罪行为尤其猖獗，引发公众对区块链安全性的质疑和对其发展前景的担忧，严重影响加密货币的价值存储功能。因此，目前迫切需要一种新的方法来更加高效而精确地识别出实施钓鱼诈骗犯罪行为的账户，从而打击区块链经济犯罪行为、保护用户的资产。

[0004] 以太坊作为下一代加密货币与去中心化应用平台，是区块链技术一次重大革新与发展。它支持通过创建智能合约发布分布式应用程序，具有成为去中心化世界虚拟机的潜质。支撑以太坊运行的以太币目前是市值排名第二的加密货币，价值超过3000亿美元。在价值居高的同时，以太坊上的网络钓鱼诈骗活动也日益猖獗。报告指出，仅在2018年，研究机构就发现以太坊上有超过2000个钓鱼诈骗账户，这些钓鱼诈骗账户从近4万人手中骗取了价值超过3600万美元的加密货币。目前，已经有一些研究者提出了对以太坊钓鱼诈骗账户的检测方法，但是还存在准确率不高的问题。因此，本发明针对这样的问题，提出了一种基于深度神经网络的简单高效的以太坊钓鱼诈骗账户检测方法。

发明内容

[0005] 针对上述问题，本发明的目的在于提供一种基于深度神经网络的以太坊钓鱼诈骗账户检测方法与装置，深度学习能自主学习到数据中的有效特征，检测结果能够明显优于传统的机器学习模型，且能够对交易进行特征分析提取进而提取钓鱼账户本身的生命周期特点，从而更为有效地鉴别钓鱼诈骗账户，检测的准确率更高。

[0006] 本发明技术方案如下：一种基于深度神经网络的以太坊钓鱼诈骗账户检测方法，包括以下步骤：

步骤1：通过网络爬虫和以太坊节点，获取账户的地址、标记和交易的相关字段，构

建以太坊钓鱼诈骗二阶子网络;从中分析并提取出以太坊钓鱼诈骗的账户交易序列特征、账户状态特征和账户交易网络特征,构建以太坊钓鱼诈骗账户数据集ETHScam;

步骤2:构建一个基于FCN-LSTM网络和BP神经网络的深度学习模型,模型命名为MFL,根据输入的以太坊钓鱼诈骗账户数据集ETHScam进行特征提取:将账户交易序列特征投入FCN(fully convolutional network,全卷积神经网络)和LSTM(long short term memory network,长短期记忆神经网络)并置的网络中提取出交易的数值特征向量和时序特征向量,将账户状态特征和账户交易网络特征投入BP神经网络学习得到统计特征向量;

步骤3:将统计特征向量、数值特征向量和时序特征向量进行拼接,然后将其输入到全连接神经网络构建的分类器中进行分类,得到账户是否为钓鱼账户的分类结果。

[0007] 进一步的,所述以太坊钓鱼诈骗二阶子网络具体包括:通过编写爬虫程序从网站Etherscan上获取标注数据;使用Bigquery服务快速查找得到需要的以太坊区块、账户和交易的统计数据;在本地服务器运行一个以太坊节点,实时与以太坊主网络同步,通过查询本地全节点同步的以太坊数据,以Etherscan标记的钓鱼诈骗账户为起点,向外枚举扩展邻点构建以太坊钓鱼诈骗二阶子网络。

[0008] 更进一步的,步骤1中,所述构建以太坊钓鱼诈骗账户数据集ETHScam具体包括:

步骤1.1:提取账户交易序列特征:选择已经被标记的账户作为钓鱼诈骗账户,随机选择出未被标记的账户作为正常账户;对于被选择的钓鱼诈骗账户和正常账户,先从子网络数据中取出其参与的所有交易记录,然后提取出每个交易中的交易时间戳、交易以太币数目、交易手续费与转账方向共计4个字段;对于交易手续费*GasPrice*,计算其与块内平均交易手续费*AvgGasPrice*的比值*GasPriceRatio*,计算公式为:

$$GasPriceRatio = \frac{GasPrice}{AvgGasPrice}$$

其中,通过Bigquery的SQL查询功能获取块内平均交易手续费*AvgGasPrice*数据;

步骤1.2:提取账户状态特征:基于Bigquery和上一步骤获取到的交易记录,计算得到账户目前的状态信息,具体为通过Bigquery查询到指定账户目前的余额;计算参与的交易数据得到账户接收和转出的以太币数目以及转出以太币数目与接收以太币数目的比值;

步骤1.3:提取账户交易网络特征:将账户参与的所有交易按照交易方向划分为转入交易和转出交易两类,统计两类交易的数目得到转入账户数目、转出账户数目以及转入转出账户数目比值;再计算两类交易的平均转账以太币数目得到平均转入以太币数目、平均转出以太币数目以及平均转入转出以太币数目比值。

[0009] 更进一步的,所述账户交易序列特征包括交易时间戳、交易以太币数目、交易方向和交易手续费比值;所述账户状态特征包括账户余额、账户涉及交易数量、接收以太币数目、转出以太币数目和以太币转出接收比值;所述账户交易网络特征包括转入账户数目、转出账户数目、转入转出账户数目比值、平均转入以太币数目、平均转出以太币数目和平均转入转出以太币数目比值。

[0010] 更进一步的,所述深度学习模型包括:

输入层:

所述输入层用于输入预处理之后得到的账户交易时间序列、账户状态特征和账户

交易网络特征；

所述输入层分三部分，第一部分将经过预处理后的账户交易时间序列TS作为输入$TS=\{TS_1,TS_2,TS_3,...,TS_n\}$；交易时间序列预处理为对原始账户交易时间序列$TS_0$使用滑动窗口采样法采样并进行归一化得到$TS$；该部分输出将会投入到特征提取层中用于提取账户交易时序序列的时序特征向量和数值特征向量；

输入层的第二部分和第三部分将账户状态特征和账户交易网络特征的统计特征向量直接并置到时序特征向量和数值特征向量之后，作为分类器的输入；

特征提取：

所述特征提取包括两大模块，分别为基于全卷积神经网络FCN的第一特征提取模块和基于LSTM的第二特征提取模块；第一特征提取模块将经过预处理后的账户交易时间序列作为输入，投入到全卷积神经网络中处理后，经过一个全局池化层得到时序变量的内部隐含特征M，M是32维的数值特征向量；第二特征提取模块用于提取时序特征，其包括8个细胞，输入层Dropout率设置为0.2，隐藏层Dropout率设置为0.5；最终输出8维的时序特征向量T；

所述账户状态特征和账户交易网络特征均为针对账户的统计特征，将此两个部分的特征向量分别进行归一化然后输入到BP神经网络中，最终得到16维的统计特征向量S；

特征拼接：

将统计特征向量S、时序特征向量T和数值特征向量M拼接得到账户特征表示向量$V=\{V_1,V_2,V_3,...,V_{56}\}$，其表示为：

$$V=T\oplus S\oplus M;$$

输出层：

将拼接起来的统计特征向量、时序特征向量和数值特征向量投入到全连接神经网络，然后通过Sigmoid函数计算得到该账户是钓鱼账户的概率，以此得到最终的分类结果P_d，其表示为：

$$p_d=sigmoid(V_E);$$

其中，V_E为最终判断账户是否为钓鱼账户的向量，经过Sigmoid函数得到预测结果；模型的优化目标为最小化交叉熵损失函数L，其表示为：

$$L=-\sum_{d\in D}[y_d log(p_d)+(1-y_d)log(1-p_d)];$$

其中，d表示样本，D表示样本数据集；y_d表示样本的真实值，p_d为样本的预测值。

[0011] 更进一步的，所述全卷积神经网络FCN包括三个时间卷积块用作特征提取器；所述卷积块包括具有多个滤波器的卷积层和多个内核，每一个卷积层都经过批量归一化；批量规范化层后接ReLU激活函数；且前两个卷积块以一个压缩和激励块结束，所有压缩和激励块的衰减率r设置为16；最后一个卷积块后接一个全局平均池化层；所述压缩和激励块带来的附加参数的总数为：

$$P=\frac{2}{r}\sum_{S=1}^{S}R_S\cdot G_S^2;$$

其中，P是附加参数的总数，r表示衰减率，S表示阶段数，G_s表示阶段S的输出特征图的数目，R_S表示阶段S的重复块数。

[0012] 一种基于深度神经网络的以太坊钓鱼诈骗账户检测装置，包括数据标注与采集模块、特征提取模块和检测模块；

所述数据标注与采集模块通过网络爬虫和以太坊节点，获取账户的地址、标记和交易的相关字段，构建以太坊钓鱼诈骗子网络；从中分析并提取出以太坊钓鱼诈骗的账户交易序列特征、账户状态特征和账户交易网络特征，构建以太坊钓鱼诈骗账户数据集ETHScam；

所述特征提取模块通过FCN和LSTM并置的网络从账户交易序列特征中提取出交易的数值特征向量和时序特征向量；通过BP神经网络从账户状态特征和账户交易网络特征中提取统计特征向量；

所述检测模块将统计特征向量、数值特征向量和时序特征向量进行拼接，再通过全连接神经网络构建的分类器进行分类。

[0013] 本发明的有益效果是：

1、本发明基于账户的交易特征与账户的状态特征，可达到97.30%的准确率。

[0014] 2、本发明提出的MFL检测模型在所有指标上均为最优；此外，基于深度学习的MFL模型的检测结果优于传统的机器学习模型，这是由于深度学习能自主学习到数据中的有效特征，而传统的机器学习需要人工进行特征提取，并且提取出所有的特征是很困难的。

[0015] 3、本发明的MFL模型在相同的网络规模下，比单纯使用LSTM和RNN网络的模型效果更好，这是因为MFL模型结合了LSTM和FCN两种网络，可以同时提取账户交易的时序特征向量和数值特征向量，再结合账户的统计特征向量，完成较大的提升。

[0016] 4、本发明提出的MFL模型能够对交易进行特征分析提取进而提取钓鱼账户的生命周期特点，从而更为有效地鉴别钓鱼诈骗账户；并且，MFL模型中融入的统计特征向量也对钓鱼账户检测结果具有一定的贡献。

[0017] 5、本发明提出的MFL模型在时序特征向量的引入、LSTM网络的使用以及时序特征向量、数值特征向量与统计特征向量的融合方面，都对最终的钓鱼账户检测结果有提升作用；因此本发明的MFL钓鱼账户检测模型在以太坊钓鱼账户检测问题上取得了较为优秀的成果。

附图说明

[0018] 图1为本发明基于深度神经网络的以太坊钓鱼诈骗账户检测方法的流程框图。

[0019] 图2为本发明MFL模型图。

[0020] 图3为特征消融结果对比。

[0021] 图4为不同时序特征感知网络的表现。

[0022] 图5为不同的检测模型和MFL模型的表现。

具体实施方式

[0023] 下面结合说明书附图和具体实施例对本发明做进一步详细说明。

[0024] 如图1所示，本发明的方法流程主要包含三个部分：数据源与数据采集、特征提取、

检测模型。

[0025] 一、数据源与数据采集:本发明所依赖的数据来自多个可信数据源,包括本地同步的以太坊网络节点和第三方服务。在这个部分中,既开发了爬虫从网站Etherscan上获取标注数据,也使用了Google Bigquery服务(https://cloud.google.com/bigquery)强大算力来进行数据库运算操作以获取以太坊账户的统计数据,还在本地服务器上同步了以太坊全节点用于进行实时的数据查找、获取账户的交易数据。基于以上的交易数据、标注数据和统计数据,构建了以太坊钓鱼诈骗子网络,并在此基础之上构建了一个以太坊钓鱼诈骗账户数据集ETHScam。

[0026] 二、账户状态特征和账户交易网络特征,并为每个账户生成对应的特征向量。本部分的核心工作在于使用FCN和LSTM网络提取账户交易序列的时序特征向量和数值特征向量。具体来说:首先,从每条交易中提取了4个特征值,并对账户参与的所有交易的特征值序列使用滑动窗口法进行采样,从中提取出共16个时间步的时序序列;然后,将得到的时间序列输入FCN和LSTM网络中,分别得到32维的数值特征向量M和8维的时序特征向量T;最后,对于账户状态特征和账户交易网络特征,将其进行拼接后送入BP(back propagation,反向传播)神经网络中学习训练,映射为32维的统计特征向量。

[0027] 三、检测模型:将“特征提取”模块中生成的统计特征向量、数值特征向量和时序特征向量拼接,然后将其输入到构建的全连接神经网络中进行分类。实验表明,相对于常见的机器学习分类器,基于全连接神经网络构建的分类器具有更高的准确率和召回率。

[0028] 1、数据源与数据采集

目前已经存在一些关于以太坊钓鱼诈骗账户检测的研究,但少有全面的、较新的公开数据集。综合使用网络爬虫、以太坊节点,基于一定的策略构建了一个以太坊钓鱼诈骗子网络,包括账户的地址、标记和交易的相关字段。基于网络,构造了一个以太坊钓鱼诈骗账户数据集ETHScam,包含44709个账户和739790条交易记录。

[0029] 1.1、数据源

(1)钓鱼诈骗账户列表与数据标注

Etherscan是一个被广泛使用的以太坊浏览器网站,经过以太坊使用者与网站开发人员的持续手工标注,该网站维护了一张以太坊钓鱼诈骗账户的列表,包含4907个钓鱼诈骗账户的地址。通过编写爬虫程序,获取了这些被标注的钓鱼诈骗账户的地址。

[0030] (2)本地以太坊节点与钓鱼诈骗子网络

在本地服务器运行了一个以太坊节点,该节点实时与以太坊主网络同步,用于在线查询以太坊数据。通过查询本地全节点同步的以太坊数据,以Etherscan标记的钓鱼诈骗账户为起点,向外枚举扩展邻点构建了以太坊钓鱼诈骗二阶子网络。该网络包含3851911个账户,18252216条交易记录,平均度为9.48。

[0031] (3)Bigquery数仓与以太坊统计数据

Bigquery是Google发布的云数据库解决方案,支持对大规模数据的实时在线查询。目前,Bigquery已经上线了以太坊全链数据库并且保持每天更新。BigQuery支持通过SQL语言进行查询,借助平台强大的计算能力可以在很短的时间内完成对以太坊全链的快速查询。可以使用Bigquery服务快速查找得到需要的以太坊区块、账户和交易的统计数据。例如,根据设计需要,查询了全链前12,000,000个区块的块内平均手续费$AvgGasPrice$值。

[0032] 1.2、数据采集

在构建的以太坊钓鱼诈骗子网络中，本实施例选择已经被标记的4709个账户作为钓鱼诈骗账户，然后随机选择出未被标记的40000个账户作为正常账户，共计44709个账户。查询这些账户参与的所有交易记录以及账户自身的状态，然后提取必要的字段，最后进行归一化操作并整理，从而得到最终的数据集ETHScam。

[0033] 首先，对于这44709个被选择的以太坊账户地址，先从子网络数据中取出该账户参与的所有交易记录，然后提取出每个交易中的交易时间戳、交易以太币数目、交易手续费与转账方向共计4个字段。对于交易手续费*GasPrice*，还会计算其与块内平均交易手续费*AvgGasPrice*的比值*GasPriceRatio*，计算公式为：

$$GasPriceRatio = \frac{GasPrice}{AvgGasPrice} \quad (1)$$

其中，通过Bigquery的SQL查询功能获取块内平均交易手续费*GasPrice*数据。由此，完成账户交易序列特征的提取。

[0034] 其次，基于Bigquery和上一步骤采集得到的交易数据，可以计算得到账户目前的状态信息。通过Bigquery可以查询到指定账户目前的余额；计算参与的交易数据可以得到账户接收和转出的以太币数目以及转出以太币数目与接收以太币数目的比值；一个账户有多少条相关的交易记录也代表了该账户参与了多少次交易。由此，完成账户状态特征的提取。

[0035] 然后，遍历账户参与的所有交易，可以得到账户的一阶邻点和这些邻点与账户的交易情况，进一步可以提取账户的交易网络特征。具体来说，本发明将账户参与的所有交易按照交易方向划分为转入交易和转出交易两类，统计两类交易的数目可以得到转入账户数目、转出账户数目以及转入转出账户数目比值。再计算两类交易的平均转账以太币数目可以得到平均转入以太币数目、平均转出以太币数目以及平均转入转出以太币数目比值。

[0036] 最后，经过整理，得到一个以太坊钓鱼诈骗数据集ETHScam，其样本分布如表1所示。

[0037]

表 1 以太坊钓鱼诈骗数据（ETHScam）描述

	数目	参与交易数目
钓鱼账户	4709	368420
正常账户	40000	371370
总计	44709	739790

2、特征提取

在特征提取模块中，对以太坊账户及其交易的相关数据进行了分析，共提出了3类(账户交易序列特征、账户状态特征、账户交易网络特征)共计15个特征，如表2所示。将账户交易序列特征投入FCN和LSTM并置的网络中提取出交易的数值特征向量和时序特征向量，将账户状态特征和账户交易网络特征投入BP神经网络学习得到统计特征向量。然后得到数值特征向量、时序特征向量和统计特征向量进行拼接，再输入全连接神经网络得到分类结

果。

表 2 特征列表

特征类别	特征名	描述
账户交易	交易时间戳	交易发生时间
	交易以太币数目	交易转移的以太币数目
	交易方向	以太币转移的方向
	交易手续费比值	交易手续费与块内平均手续费的比值
账户状态	账户余额	账户目前持有的以太币数目
	账户涉及交易数量	账户参与了多少次交易
	接收以太币数目	账户已经接收的以太币总数目
	转出以太币数目	账户已经转出的以太币总数目
	以太币转出接收比值	账户转出以太币数目与接收以太币数目的比值
账户一阶交易网络特征	转入账户数目	向本账户转入以太币的账户数目
	转出账户数目	本账户向外转出以太币的账户数目
	转入转出账户数目比值	转入账户数目与转出账户数目的比值
	平均转入以太币数目	转入交易平均向本账户转入的以太币数目
	平均转出以太币数目	本账户平均向外转出的以太币数目
	平均转入转出以太币数目比值	平均转入以太币数目与平均转出以太币数目的比值

[0038]

2.1、账户交易特征

账户交易特征主要是通过构建一组时序向量来描述账户所参与所有交易的时序特征。

[0039] 依据现有的研究成果并参考对传统钓鱼行为的研究，以太坊上发生的钓鱼行为大致可以分为三个阶段：发展期、泛滥期和结束期。

[0040] 发展期：攻击者构造钓鱼欺诈信息并通过多种渠道将其散播到各个平台、社区和

网络上，但是钓鱼欺诈信息需要一定的时间来进行广泛传播。因此，该时期内受骗的人数较少，钓鱼账户参与的交易多具有少量、小额、低频的特点。

[0041] 泛滥期：经过一定时间，钓鱼欺诈信息已经得到了充分的传播，并在各个平台社区内成功欺骗到了大量的受害者，受害者被诱导向钓鱼账户发送以太币。在这个时期内，钓鱼账户会参与大量、足额且高频的交易中。

[0042] 结束期：随着受害者人数的上升，不少人逐渐意识到自己被钓鱼攻击。同时，大量的钓鱼欺诈信息还可能会引起平台方和社区成员的警觉，并促使他们发布警示信息。即使是少量的警示信息也会对钓鱼欺诈信息的有效性造成巨大的破坏，钓鱼行为也会因此遭到猛烈打击而迅速进入结束期。在该时期内，钓鱼欺诈信息很难取得人们的信任，鲜有用户发起向钓鱼账户转账的交易。与此同时，钓鱼者可能会开始将账户余额向外转移变现，从而牟利。

[0043] 以上的钓鱼诈骗账户的生命周期特点可以作为检测钓鱼账户的重要依据。

[0044] 采用滑动窗口采样法，对于每个账户参与的所有交易记录提取出16个时间步的特征序列，每个时间步内部包含4个特征值。借助FCN和LSTM并置的神经网络来提取时序序列特征值的分布规律以及在时间上的演变规律，以此为重要依据判别账户是否为钓鱼诈骗账户。

[0045] (1)交易时间戳：交易时间戳描述了交易发起的时间，通过时间戳可以描述了账户参与的所有交易在时间上的分布规律。钓鱼账户往往在发展期涉及少量低频交易，在爆发期涉及大量、高频的交易，在结束期涉及少量的交易。

[0046] (2)交易以太币数目：交易以太币数目指在一次交易中，由交易发起方向接收方转移的以太币数量。以太币的最小计量单位为*wei*，一个以太币等于10^{18}*wei*。这里以以太币的最小计量单位*wei*来记录每次交易转移的以太币数量。

[0047] (3)交易方向：对于钓鱼账户，其所参与交易的方向可以作为判断交易类型的重要依据。向钓鱼账户转入以太币的交易往往是受害者发起的支付资金的交易，方向特征值记作-1。由钓鱼账户向外转出以太币的交易往往是钓鱼者为了转移资产而发起的，方向特征值记作1。

[0048] (4)交易手续费比值：随着以太坊平台生态的日益完善，在以太坊上运行的应用、进行的交易越来越多，这不可避免地导致以太坊网络产生了性能瓶颈，大量的交易缓存在网络中等待矿工将其打包记入链上。完成一次交易需要一定的计算量，交易中的*GasPrice*字段描述了交易发起者愿意为单位计算量所支付的手续费用。手续费越高，矿工完成交易所得到收益越大，所以矿工会优先完成手续费高的交易。为了确保受害者尽快地成功向钓鱼账户转入以太币并将交易记录写入区块链上，钓鱼者会诱导受害者设置较高的*GasPrice*值。但是，以太币的价值一直处在波动变化当中，相应的平均手续费也会发生变化。为了解决这个问题，本实施例计算了当前交易的手续费与同一区块内平均手续费的比值*GasPriceRatio*。较高的*GasPriceRatio*数值可以说明交易的发起者迫切希望本次交易能够尽快被记入链上，可以作为判断钓鱼账户重要依据。

[0049] 2.2、账户状态特征

以太坊上普通用户的状态往往具有很强的同质性。因为多数账户为散户，这些账户具体表现为持有少量以太币、参与少量交易、非活跃状态，这与钓鱼诈骗账户的状态区别

很大。钓鱼诈骗账户往往涉及较多的交易、持有或者曾经较多的以太币、并且会在一定时期内活动频繁，因为该时期很可能是账户吸收、转移诈骗资金的时期。

[0050] (1)账户余额：账户目前持有的以太币余额。钓鱼账户吸收了大量的诈骗资金从而保留有较多的以太币余额。

[0051] (2)账户参与交易数量：以太坊上的账户以散户居多，散户交易慎重，参与的交易总数较少，因此账户参与交易的总数可以作为判断钓鱼账户的重要依据。

[0052] (3)账户接收、转出以太币数目及其比值：一次成功的钓鱼欺诈行为往往会吸收到许多的以太币，数目会超过普通账户持有的以太币数目。接收大量的以太币是钓鱼账户的重要特征。在吸收得到大量的以太币之后，钓鱼者需要经过资金转移将以太币兑换为其他加密货币或者法币获取实际的经济收益，完成这个步骤需要先将钓鱼账户的以太币转移到一个或者多个中间账户。使用特定的账户用于多次钓鱼欺诈并非长久之计，钓鱼者往往会在钓鱼行为结束之后将账户的以太币全部转出变现，一个账户不会重复使用。因此，将吸收到的以太币全部转出是钓鱼账户的重要标志。

[0053] 2.3、账户交易网络特征

账户交易网络由账户自身和与它发生交易的账户加上账户之间的交易记录组成。向钓鱼诈骗账户发起转账的账户往往是受害者创建的、用以支付资金的账户，钓鱼诈骗账户资金转出的目标账户往往是犯罪分子用于洗钱变现的账户。

[0054] (1)转入转出账户数目及比值：为了获得最大的利益，钓鱼者会引诱尽可能多的受害者发起转账，因此向钓鱼账户转入以太币的账户数目较多。同时，为了逃避资金追溯，钓鱼者还会创建少量的中间账户用于洗钱变现。

[0055] (2)平均转入转出以太币数目及比值：为了使得受害者能够完成以太币转账，钓鱼者会设置一个合适的金额，该金额不会过大超过受害者的经济能力，也不会过小从而降低钓鱼者的收益。同时，为了加快洗钱变现过程，钓鱼者会通过少量大额的交易转出钓鱼账户的以太币。因此，计算转入以太币的平均数目和转出以太币的平均数目以及两者之间的比值将有助于判定钓鱼诈骗账户。

[0056] 3、检测模型

本发明设计了一个基于多特征融合和FCN-LSTM的MFL(namely Multivariate FCN-LSTM model)深度学习模型来进行以太坊钓鱼诈骗账户检测。MFL模型基于FCN-LSTM模型，并融合了squeeze-and-excitation block机制使得FCN-LSTM可以处理多元变量的时间序列。同时结合了账户状态特征与账户交易网络特征，能够有效地检测以太坊中的钓鱼诈骗账户。MLF模型结构如图2所示。

[0057] 3.1、输入层

本发明提出的MFL模型的输入主要分为三个部分：预处理之后得到的账户交易时间序列、账户状态特征和账户交易网络特征。

[0058] 其中，账户交易时间序列步长为16，每一步包含4个特征值。账户状态特征共计5个，账户交易网络特征共计6个。模型最终的输出为账户是钓鱼诈骗账户的概率。

[0059] 如图2中“输入层”所示，输入层的第一部分将经过预处理后的账户交易时间序列*TS*作为输入$TS=\{TS_1,TS_2,TS_3,...,TS_n\}$。交易时间序列预处理为对原始账户交易时间

序列TS_0使用滑动窗口采样法采样并进行归一化得到TS。经过分析，以太坊上面的多数账户所参与的交易数量不超过16，因此这里设置n=16。对于原始交易时间序列少于16的，将在末尾填充0。对于超过16的，将进行多次采样，每次采样之后将窗口向后移动4步。该部分输出将会投入到“特征提取”模块中用于提取账户交易时序序列的时序特征向量和数值特征向量。

[0060] 输入层的第二部分和第三部分的处理较为类似，因为两者都属于统计特征。因此，将这两个部分的特征值分别进行归一化操作之后输入到BP神经网络学习特征，然后将输出的统计特征向量直接并置到时序特征向量和数值特征向量之后，作为分类器的输入。

[0061] 3.2、特征提取

MFL模型的时序序列特征提取分为两大模块，分别是基于完全卷积神经网络FCN的特征提取模块和基于LSTM的特征提取模块，如图2所示。

[0062] 全卷积神经网络包含三个时间卷积块用作特征提取器。卷积块包含具有多个滤波器(大小分别为32、32和32)的卷积层和多个内核(大小分别为8、5和3)。每一个卷积层都经过批量归一化，归一化动量为0.99，ε为0.001。批量规范化层后接ReLU激活函数。此外，前两个卷积块以一个squeeze-and-excite(压缩和激励)块结束，这使该模型区别于传统的FCN-LSTM。对于所有压缩和激励块，将衰减率设置为16。最后一个卷积层后接一个全局平均池化层。

[0063] FCN块可以自适应地重新校准输入特征映射，而挤压和激励块是FCN块的一个补充。由于衰减率设置r为16，学习这些自注意力图所需的参数数量有所减少，因此整体模型大小仅增加3-10%。其计算方法如下：

$$P=\frac{2}{r}\sum_{S=1}^{S}R_S\cdot G_S^2 \quad (2)$$

其中，P是附加参数的总数，r表示衰减率，S表示阶段数，G_s表示阶段S的输出特征图的数目，R_S表示阶段S的重复块数。

[0064] 压缩和激励块对于增强多变量数据集的性能至关重要，因为并非所有特征映射都会对后续层产生相同程度的影响。这种特征映射的自适应重新校准可以看作是对先前层的输出特征映射学习的自我关注的一种形式。与传统的FCN-LSTM相比，这种滤波器映射的自适应重缩放对于多变量的FCN-LSTM模型的性能改进至关重要，因为它将学习到的自我关注纳入到每个时间步多个变量之间的相互关系中，而传统的FCN-LSTM不具备这种能力。

[0065] 具体来说，FCN将经过预处理后的账户交易时间序列TS作为输入，然后投入到全卷积神经网络中，最后经过一个全局池化层得到时序变量的内部隐含特征M，M是32维的数值特征向量。

[0066] 另外，使用LSTM来提取时序特征。因为多元时间序列是经过滑动窗口采样法得到的，不同的时间步之间存在真实的先后顺序，所以在这里直接投入LSTM网络。LSTM网络包含8个细胞，输入层Dropout率设置为0.2，隐藏层Dropout率设置为0.5。最终输出8维的时序特征向量T。

[0067] 对于输入层的账户状态特征和账户交易网络特征，因为两者都是，所以对这两个部分的特征向量分别进行归一化然后输入到BP神经网络中，最终得到16维的统计特征向量

S。

[0068] 3.3、特征拼接

本发明借助Keras拼接技术将得到“特征提取”层输出的三类特征向量融合，从而获得最终的账户特征表示向量V，并将其输入至全连接神经网络得到账户分类结果。

[0069] 统计特征向量作为钓鱼账户检测中账户的全局属性，其能够从全局的角度区分钓鱼账户与正常账户。但是统计特征向量仅仅对账户属性进行了统计，无法获得账户所参与交易的时序特征。因此，本发明将统计特征向量与交易时序特征向量以及交易的数值特征向量相结合，能够扩充钓鱼诈骗账户检测的特征空间，也能在更大程度上描述数据在特征空间中的分布，从而提高网络的分类性能。

[0070] 如图2中“特征拼接”所示，将统计特征向量S、时序特征向量T和数值特征向量M拼接得到账户特征表示向量$V=\{V_1,V_2,V_3,\ldots,V_{56}\}$，其表示为：

$$V=T\oplus S\oplus M \tag{3}$$

3.4、输出层

最后，将拼接起来的统计特征向量、时序特征向量和数值特征向量投入到全连接神经网络，然后通过Sigmoid函数计算得到该账户是钓鱼账户的概率，以此得到最终的分类结果P_d，其表示为：

$$p_d=sigmoid(V_E) \tag{4}$$

其中，V_E为最终判断账户是否为钓鱼账户的向量，经过Sigmoid函数得到预测结果。

[0071] 模型的优化目标为最小化交叉熵损失函数L，其表示为：

$$L=-\sum_{d\in D}[y_d log(p_d)+(1-y_d)log(1-p_d)] \tag{5}$$

其中，d表示样本，D表示样本数据集，y_d表示样本的真实值，p_d为样本的预测值。在二分类的结果中，0表示正常账户，1表示钓鱼账户。

[0072] MFL模型的部分参数设置如表3所示。

表 3 MFL 模型部分参数列表

[0073]

模型参数	参数值
Batch Size	128
Epoch	100
Early Stopping	Min_delta=0.001, Patience=20
Optimizer	Adam(clipvalue=5.0)
Loss	binary_crossentropy
LSTM_Cell_number	8
LSTM_Dropout	0.2
LSTM_Recurrent_Dropout	0.5
FCN_kernel_size	8,5,3
FCN_filter_number	32,32,32
FCN_activiation	ReLU
BP_units	16
Dense_units	32
Dense_Dropout	0.5

3.5、模型训练流程

本发明借鉴相关研究在构建数据集工作方面的思路，构建了以太坊钓鱼诈骗数据集ETHScam，共包含4907个已知钓鱼诈骗账户和40000个正常账户。然后构建了一个基于FCN-LSTM的模型MFL。MFL可以分析账户的交易时序序列提取交易的时序特征向量与数值特征向量，同时借助BP神经网络提取账户的统计特征向量，然后将向量并置投入全连接神经网络，综合分析判断账户类别。在训练过程中，使用了早停机制，设置学习率为0.001，提前退出阈值设置为0.01，Batch Size为128，最多训练200个epoch。

[0074] 4、实验

设计了三个实验来评估MFL模型的以太坊钓鱼诈骗账户检测效果。所有实验在搭载Nvidia RTX 2080 8G的服务器环境下进行，数据集为本项目收集的ETHScam数据集，共包含4907个钓鱼诈骗账户和40000正常账户。实验中划分数据集的90%作为训练集，10%作为测试集。每次实验取10折交叉验证结果的平均值作为最终结果。

[0075] 4.1、评估统计特征的有效性

为了评估本发明提出的三种类别的统计特征(账户交易序列特征、账户状态特征、账户交易网络特征)在提出的MFL检测模型中的贡献，进行了特征消融实验，在全特征集与

四个特征子集上进行了实验，特征集合如表4所示。

表 4 特征集描述

[0076]

特征集名称	包含的特征类别
F	账户交易特征、账户状态特征、账户一阶交易网络特征
F\Transactions	账户状态特征、账户一阶交易网络特征
F\State	账户交易特征、账户一阶交易网络特征
F\Network	账户交易特征、账户状态特征

实验结果如图3和表5所示。可以看到，统计特征的全特征集表现最佳，具体如表5首行所示，说明本发明提取的三种类型特征能够从多个角度提升钓鱼诈骗账户的检测效果。除此以外，MFL模型在使用F\Transactions特征子集时表现最差，说明具有账户交易特征对钓鱼诈骗账户检测具有重要的意义，这也与以太坊中实际发生的钓鱼诈骗犯罪行为的真实情境相符。

[0077] 使用F\State和F\Network特征子集的效果相近且与使用特征全集*F*的效果差距最小，这表明账户状态特征和账户交易网络特征对模型检测钓鱼诈骗账户将测均有一定的贡献度但是两者相关性很强。分析可能的原因是由于账户状态特征与账户交易网络特征都是部分基于统计和计算账户所参与的所有交易数据得到的，因此两者共享了部分隐含特征，这使得统计特征向量在钓鱼诈骗账户检测中没有发挥出最佳的效果。

表 5 特征消融结果对比

[0078]

特征集名称	准确率	精确率	召回率	F 值
F	0.9730	0.9813	0.9759	0.9786
F\Transactions	0.8781	0.8915	0.9189	0.9050
F\State	0.9627	0.9796	0.9610	0.9702
F\Network	0.9626	0.9789	0.9615	0.9702

4.2、交易记录时序特征提取效果

MFL模型的时序特征提取器同时使用FCN和LSTM来提取账户所参与交易的特征，其中FCN用于提取交易的数值特征向量，LSTM用于提取交易的时序特征向量。为了评估LSTM在提取时序特征向量方面的有效性，设计实验对比了常用的时序特征感知网络BiLSTM(Bidirectional Long-Short Term Memory，双向长短期记忆网络)和GRU(Gated Recurrent Network，门控神经网络)。在试验过程中，分别使用FCN-LSTM、FCN-GRU和FCN-BiLSTM网络结构作为账户交易时序特征提取器，其余的结构保持不变。同时添加一组无时序特征提取器和一组无FCN网络的实验模型，用0填充缺失的FCN或LSTM输出向量。

[0079] (1)GRU:GRU网络包含8个细胞，输入层dropout率设置为0.2，隐藏层dropout率设置为0.5。网络接收的数据为16个时间步、每个时间步包含4个特征的时间序列，输出的时序

特征向量的维度为8。

[0080] (2)LSTM:LSTM网络包含8个细胞,输入层dropout率设置为0.2,隐藏层dropout率设置为0.5。网络接收的数据为16个时间步、每个时间步包含4个特征的时间序列,输出的时序特征向量的维度为8。

[0081] (3)BiLSTM:BiLSTM网络包含8个细胞,输入层dropout率设置为0.2,隐藏层dropout率设置为0.5。网络接收的数据为16个时间步、每个时间步包含4个特征的时间序列,输出的时序特征向量的维度为8。

[0082] (4)FCN:FCN网络包含3个卷积层,每个卷积层都含有32个滤波器,滤波器内核大小依次为8,5,3。

[0083] 四种不同时序特征提取器的描述如表6所示。

表 6 不同时序特征提取器的描述

[0084]

时序特征提取器	输出维度	参数数量
FCN+GRU	40	10480
FCN+LSTM	40	10560
FCN+BiLSTM	40	10976
LSTM	8	608
FCN	32	9952

实验结果如图4和表7所示。总体来说,在钓鱼账户检测方面,使用了LSTM或BiLSTM的模型表现优于仅使用FCN的模型,这是由于LSTM网络能够基于时序上下文提取交易的时序特征向量,而上下文无关的模型对所有的交易信息只进行数值分析而不分析交易在时间上的内在联系。

[0085] 此外,MFL模型使用LSTM取得了较使用GRU的模型而言更好的效果,这是由于GRU相对于LSTM只使用了2个门控开关,包含的参数数量更少,效果很难超过LSTM。另外,LSTM具备实现长期依赖的能力,能够更好地感知距离当前时间步较远的交易,在长时间的上下文提取能力上有更为明显的优势。

表 7 不同时序特征感知网络的表现

[0086]

时序特征提取网络	准确率	精确率	召回率	F 值
FCN+LSTM	0.9730	0.9817	0.9761	0.9786
FCN+GRU	0.9712	0.9795	0.9747	0.9771
FCN+BiLSTM	0.9720	0.9813	0.9738	0.9777
FCN	0.9712	0.9782	0.9759	0.9771
RNN	0.9351	0.9745	0.9214	0.9472

4.3、评估提出的检测模型的效果：

为了证明本发明提出的MFL模型在以太坊钓鱼诈骗账户检测中有明显的优势，挑选了包括传统机器学习和深度学习在内的常用的钓鱼账户检测模型进行了实验，包括SVM(Support Vector Machine，支持向量机)、BiLSTM、DT(Decision Tree，决策树)以及RF(Random Forest，随机森林)模型，并且分别在准确率、精确率、召回率、F1得分等指标上进行了对比。

[0087] 实验结果如图5和表8所示。可以看到，本发明提出的MFL检测模型在构建的ETHScam数据集上取得了0.9786的F1得分，且在所有指标上均为最优。此外，基于深度学习的MFL模型的检测结果优于传统的机器学习模型，这是由于深度学习能自主学习到数据中的有效特征，而传统的机器学习需要人工进行特征提取，并且提取出所有的特征是很困难的。

[0088] 并且，本发明的MFL模型在相同的网络规模下，比单纯使用LSTM和RNN网络的模型效果更好，这是因为MFL模型结合了LSTM和FCN两种网络，可以同时提取账户交易的时序特征向量和数值特征向量，再结合账户的统计特征向量，完成较大的提升。

[0089] 最后，对比在现有研究常用的模型，即SVM、DT和RF，可以发现MFL模型更适用于钓鱼账户检测问题，这是因为钓鱼账户本身的生命周期可以被总结抽象，且每个生命周期内都表现出了特定的活动特点。而本发明提出的MFL模型能够对交易进行特征分析提取进而提取这些周期特点，从而更为有效地鉴别钓鱼诈骗账户。并且，MFL模型中融入的统计特征向量也对钓鱼账户检测结果具有一定的贡献。

表 8 不同的检测模型和 MFL 模型的表现

[0090]

模型	准确率	精确率	召回率	F 值
MFL	0.9730	0.9813	0.9759	0.9786
LSTM	0.9325	0.9453	0.9486	0.9470
RNN	0.9332	0.9425	0.9530	0.9477
SVM	0.235	0.9517	0.9263	0.9388
DT	0.9073	0.9266	0.9271	0.9269
RF	0.9314	0.9370	0.9560	0.9464

综上所述，本发明提出的MFL模型在时序特征向量的引入、LSTM网络的使用以及时序特征向量、数值特征向量与统计特征向量的融合方面，都对最终的钓鱼账户检测结果有一定的提升作用。因此，本发明的MFL钓鱼账户检测模型在以太坊钓鱼账户检测问题上取得了优秀的成果。

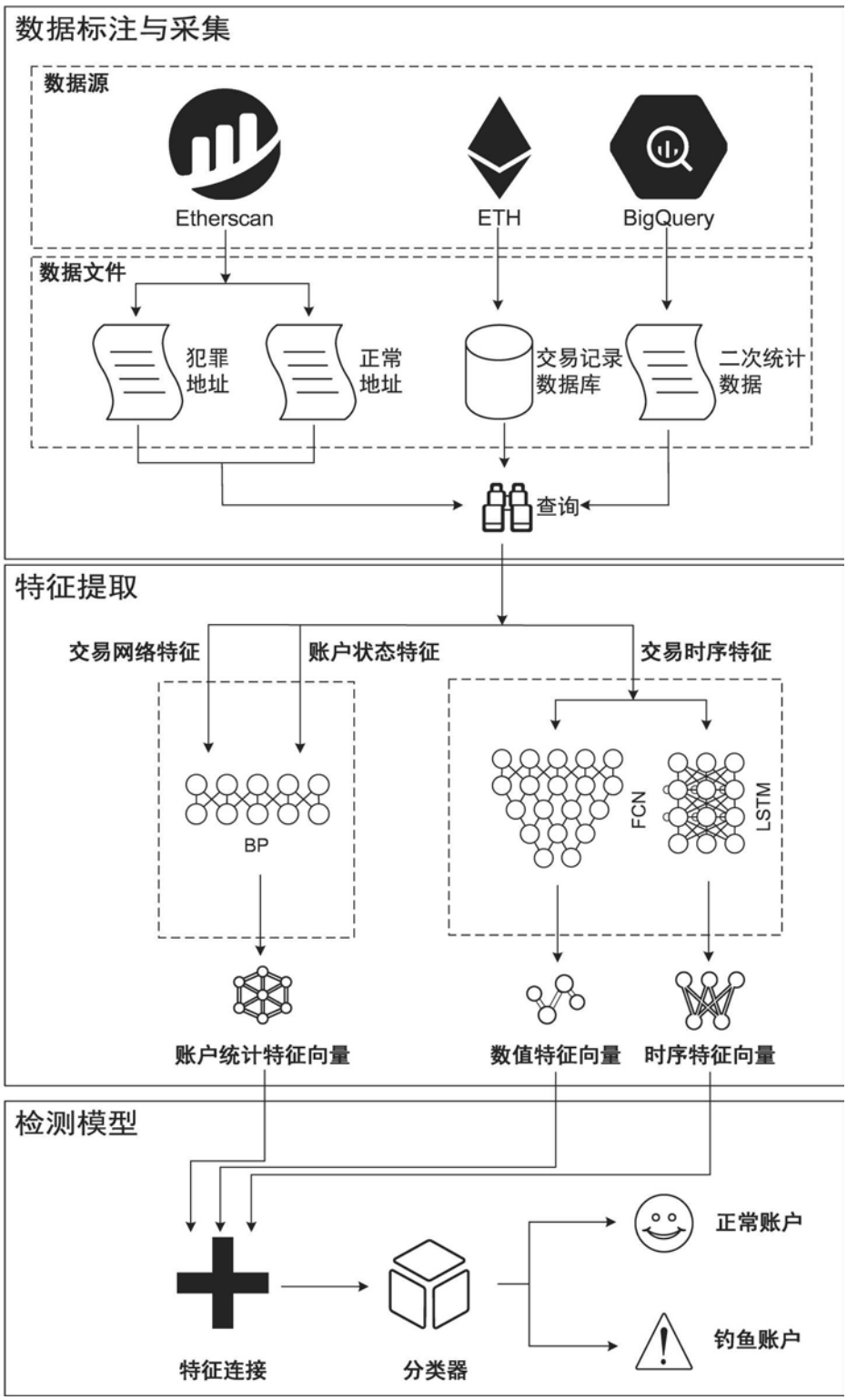

数据标注与采集
数据源
Etherscan
ETH
BigQuery
数据文件
犯罪地址
正常地址
交易记录数据库
二次统计数据
查询
特征提取
交易网络特征
账户状态特征
交易时序特征
BP
FCN
LSTM
账户统计特征向量
数值特征向量
时序特征向量
检测模型
特征连接
分类器
正常账户
钓鱼账户

图1

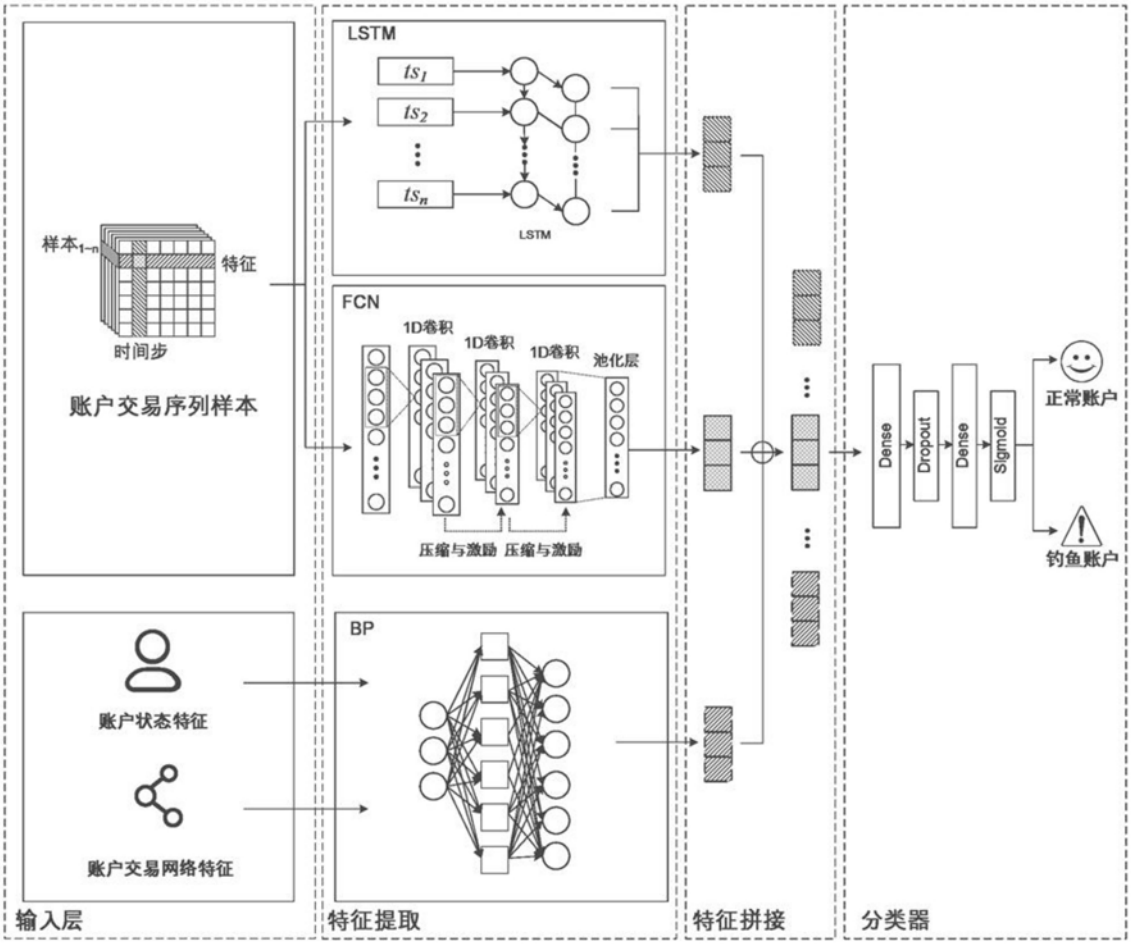

LSTM
ts_1
ts_2
ts_n
LSTM
样本1-n
特征
时间步
账户交易序列样本
FCN
1D卷积
1D卷积
1D卷积
池化层
压缩与激励
压缩与激励
BP
账户状态特征
账户交易网络特征
Dense
Dropout
Dense
Sigmoid
正常账户
钓鱼账户
输入层
特征提取
特征拼接
分类器

图2

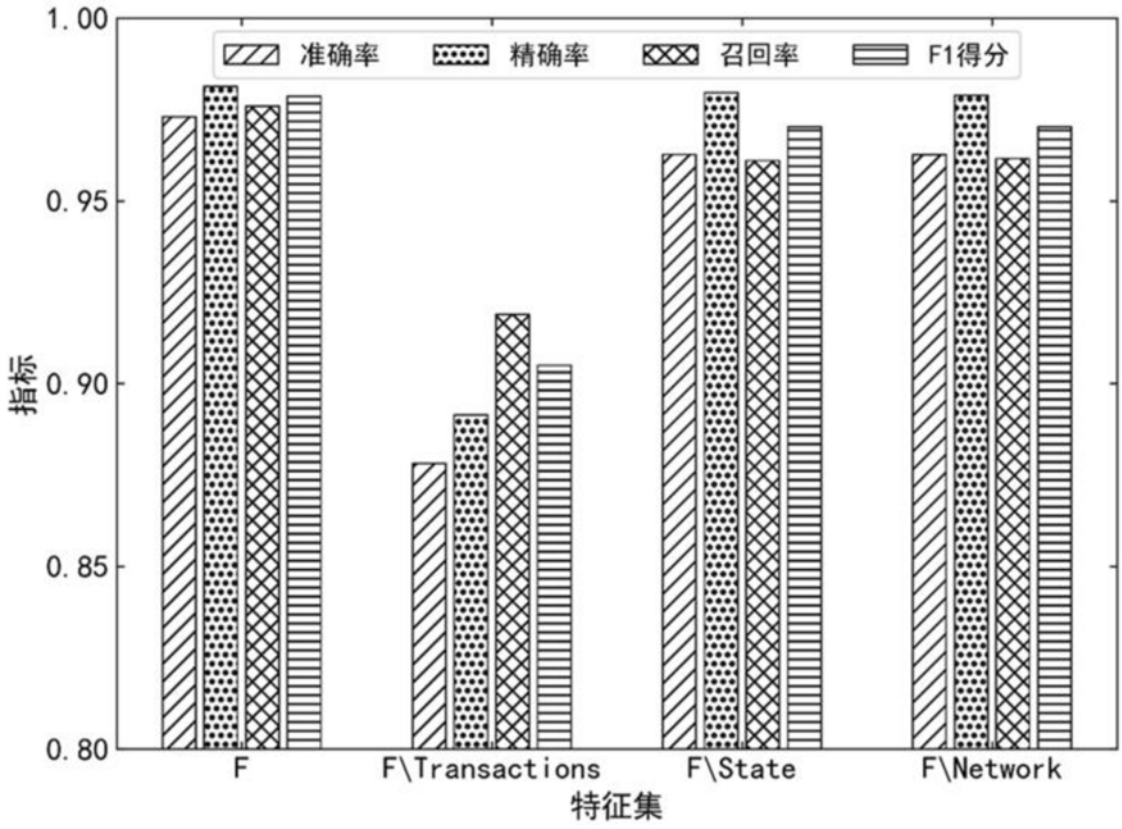

1.00
0.95
0.90
0.85
0.80
指标
准确率
精确率
召回率
F1得分
F
F\Transactions
F\State
F\Network
特征集

图3

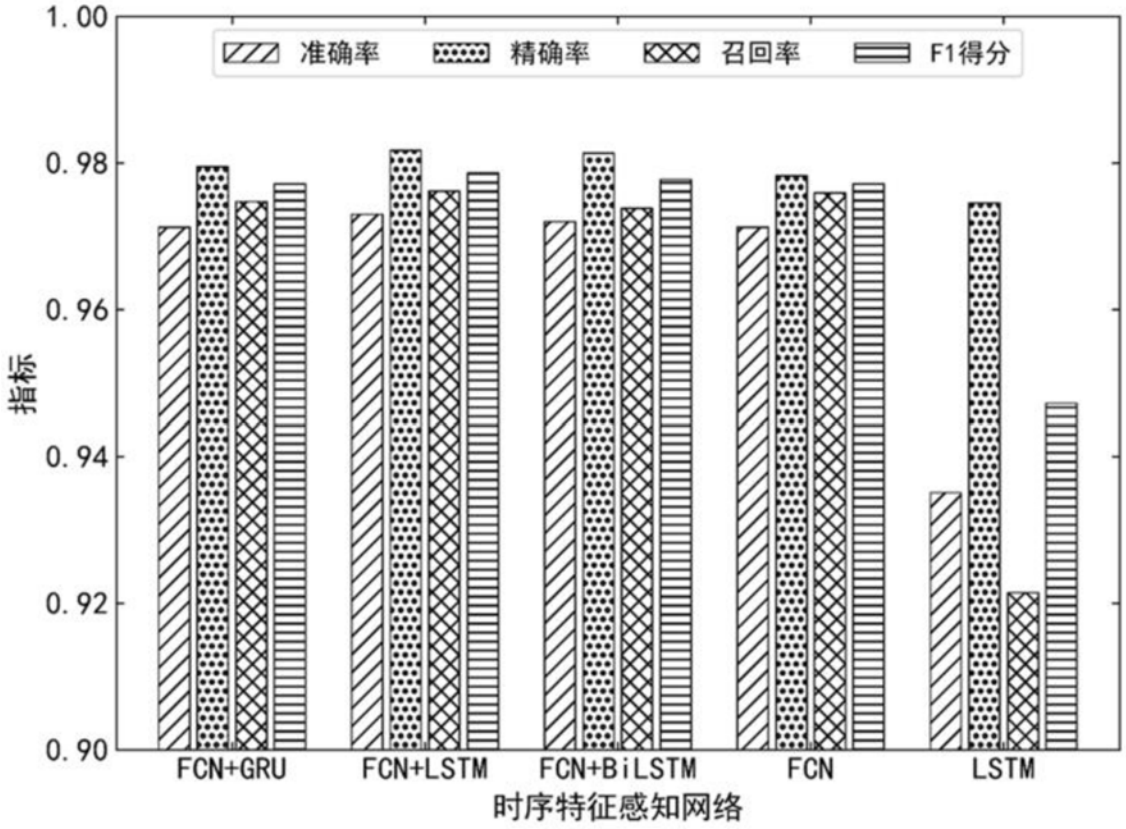

1.00
0.98
0.96
0.94
0.92
0.90
指标
准确率
精确率
召回率
F1得分
FCN+GRU
FCN+LSTM
FCN+BiLSTM
FCN
LSTM
时序特征感知网络

图4

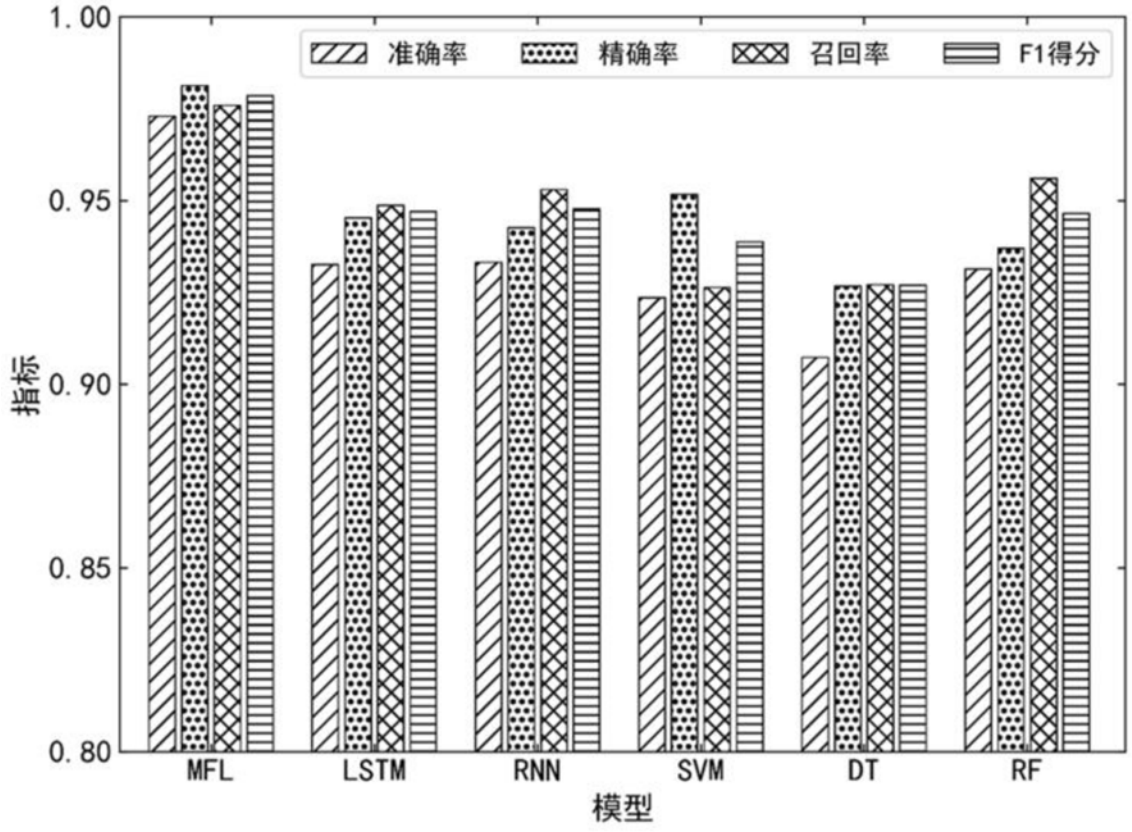


图5