(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 113191775 A
(43)申请公布日 2021.07.30

(21)申请号 202110434860.6

(22)申请日 2021.04.22

(71)申请人 深圳前海移联科技有限公司
地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室

(72)发明人 罗少龙 袁妙 胥勇

(74)专利代理机构 深圳市中融创智专利代理事务所(普通合伙) 44589
代理人 叶垚平

(51)Int.Cl.
G06Q 20/40(2012.01)
G06K 9/62(2006.01)

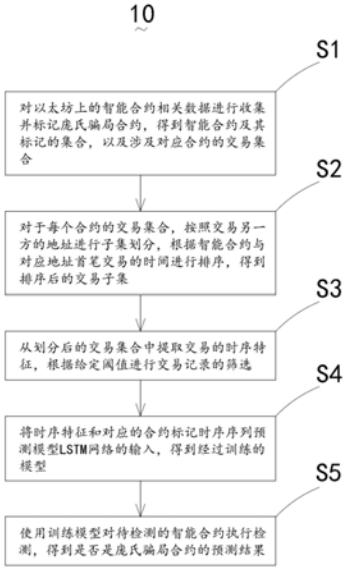
权利要求书2页 说明书8页 附图5页

(54)发明名称

基于以太坊上交易时序信息的庞氏骗局智能合约检测方法

(57)摘要

本发明公开了基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,包括以下步骤,对以太坊上的智能合约相关数据进行收集并标记庞氏骗局合约,得到智能合约及其标记的集合,以及涉及对应合约的交易集合;对于每个合约的交易集合,按照交易另一方的地址进行子集划分,根据智能合约与对应地址首笔交易的时间进行排序,得到排序后的交易子集;从划分后的交易集合中提取交易的时序特征,根据给定阈值进行交易记录的筛选;将时序特征和对应的合约标记时序序列预测模型LSTM网络的输入,得到经过训练的模型;本发明的基于以太坊上交易时序信息的庞氏骗局智能合约检测方法具有可扩展性、可解释性等优点。



CN 113191775 A

1.基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,包括以下步骤:

步骤S1:对以太坊上的智能合约相关数据进行收集并标记庞氏骗局合约,得到智能合约及其标记的集合,以及涉及对应合约的交易集合;

步骤S2:对于每个合约的交易集合,按照交易另一方的地址进行子集划分,根据智能合约与对应地址首笔交易的时间进行排序,得到排序后的交易子集;

步骤S3:从划分后的交易集合中提取交易的时序特征,根据给定阈值进行交易记录的筛选;

步骤S4:将时序特征和对应的合约标记时序序列预测模型LSTM网络的输入,得到经过训练的模型;

步骤S5:使用训练模型对待检测的智能合约执行检测,得到是否是庞氏骗局合约的预测结果;

其中,智能合约标记的集合为$\{(c_1,label_1),(c_2,label_2),(c_3,label_3),\cdots,(c_n,label_n)\}$,对应合约的交易集合为$\{T_1,T_2,T_3,\cdots,T_n\}$,其中$T_i$为第i个智能合约对应的交易集合。

2.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,所述步骤S1包括:

步骤S11:人工阅读合约源代码,根据合约是否实现庞氏骗局逻辑进行标记;

步骤S12:通过交叉检查减少错误标记率,判断智能合约是否实现了庞氏骗局逻辑;

步骤S13:若智能合约实现了庞氏骗局逻辑,则将此样本标记为正样本;

步骤S14:若智能合约未实现庞氏骗局逻辑,则将此样本标记为负样本。

3.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,所述步骤S2包括:

步骤S21:对于每个智能合约,以与智能合约发生交易的账户地址为分类标准,将交易记录分为k个子集;

步骤S22:按照每个账户与智能合约发生第一笔交易的时间,对k个子集进行排序,得到划分后的交易记录;

步骤S23:将所有智能合约的交易记录集合起来,得到总的交易记录划分;

其中,k为与智能合约发生交易的账户数量,划分后的交易记录为$\{T_{i1},T_{i2},\cdots,T_{ik}\}$,总的交易记录划分$\{T_{11},T_{12},\cdots,T_{n1},T_{n2},\cdots,T_{nk}\}$。

4.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,所述步骤S3包括:从“使用后来投资者的投资作为前面投资者的利润”这一庞氏骗局的固有交易模式出发,对智能合约与每个交易账户的交易记录子集$\{T_{i1},T_{i2},\cdots,T_{ik}\}$,提取包括以下特征:

转入交易的数量,交易源地址为对应交易账户,目的地址为该智能合约的交易数量;

转出交易的数量,交易源地址为该智能合约,目的地址为对应交易账户的交易数量;

智能合约与该账户总交易数量;

第一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生第一笔交易的时间;

最后一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生最后一笔交易的时间;

发起第一笔交易时合约的余额;

与该账户的第一笔交易是转入交易还是转出交易;

交易总回报,转出交易的总金额减去转入交易的总金额;

智能合约与该账户交易时间周期,最后一笔交易时间与第一笔交易时间的差值;

对每个智能合约,通过上述步骤提取得到大小为(k,9)的时序特征矩阵X_i,对数据集中所有智能合约执行上述特征提取步骤,得到n个时序特征矩阵。

5.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,所述步骤S4包括:

步骤S41:判断合约发生交易的地址数量k与k_{min}、k_{max}的大小关系;

步骤S42:如果与合约发生交易的地址数量k满足$k<k_{min}$,则认为合约没有足够的交易信息来帮助模型检测庞氏骗局合约,将这些合约排除在外;

步骤S43:如果与合约发生交易的地址数量k满足$k_{min}<k<k_{max}$,则在大小为(k,9)的时序特征矩阵后补$k_{max}-k$行,9列的0元素作为填充,得到大小为(k_{max},9)的时序特征矩阵;

步骤S44:如果与合约发生交易的地址数量k满足$k>k_{max}$,则从时序特征矩阵中均匀采样k_{max}行数据,得到大小为(k_{max},9)的时序特征矩阵;

步骤S45:综合得到大小为(N,k_{max},9)的特征矩阵;

在(N,k_{max},9)的特征矩阵中,N代表每个智能合约,k_{max}代表与该智能合约进行交易的账户,9代表代表该账户与该智能合约交易的特征。

6.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,所述步骤S5包括:将提取到的智能合约交易时序特征与对应标记输入时序分类模型进行训练时,对LSTM网络进行参数配置和调整,在数据集上进行训练,测试以及验证。

7.根据权利要求1所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,其特征在于,对于待分类的智能合约,所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法还包括:

步骤S100:收集待分类智能合约的交易记录,对交易记录进行时序特征提取得到大小为(k,9)的时序特征矩阵;

步骤S110:判断智能合约交易的账户数量k与k_{min}、k_{max}的大小关系;

步骤S120:若与该智能合约交易的账户数量k满足$k<k_{max}$,则在时序特征矩阵后补0得到大小为(k_{max},9)的时序特征矩阵;

步骤S130:若与该智能合约交易的账户数量k满足$k<k_{min}$,则提示预测结果可信度不高;

步骤S140:若与该智能合约交易的账户数量k满足$k>k_{max}$,则对矩阵的行进行均匀抽样,得到大小为(k_{max},9)的时序特征矩阵;

步骤S150:将得到的时序特征输入训练好的LSTM网络,得到预测结果。

基于以太坊上交易时序信息的庞氏骗局智能合约检测方法

技术领域

[0001] 本发明涉及区块链技术领域,具体涉及基于以太坊上交易时序信息的庞氏骗局智能合约检测方法。

背景技术

[0002] 区块链是一种分布式账本技术,利用密码学技术,分布式网络和智能合约技术,区块链为用户提供了一个互信的多方合作环境。其中,智能合约为区块链与各个特定领域的结合提供了基础。智能合约本质上是一段部署在区块链上的代码,因此用户可以通过智能合约实现复杂的业务逻辑,并保证智能合约的运行享有区块链上数据不可篡改,公开访问,匿名性和可溯源性的特点。以太坊是目前具有最多智能合约的区块链平台,大量的智能合约在以太坊上运行和执行交易。然而以太坊不要求智能合约提供合约的源代码,对于这些智能合约,用户难以判断合约的执行逻辑是否是符合期望,导致一系列恶意的智能合约混淆在运行正常逻辑的智能合约当中。其中,实现庞氏骗局逻辑的智能合约已经带来了巨大的经济损失,对以太坊的生态健康造成危害。

[0003] 时序序列预测模型:长短时记忆网络(Long Short-Term Memory,LSTM)是一种被广泛应用的,基于神经网络技术的机器学习模型,该模型在时序序列的预测和分类问题上取得了优秀的成果。在庞氏骗局合约检测方案中,将从庞氏骗局合约交易记录中提取到的与交易时间相关的信息作为神经网络的训练输入,将合约是否为庞氏骗局合约作为训练标签,通过反向传播方式进行迭代训练可以得到一个分类模型,该模型可以对新的输入进行分类,预测该输入对应的智能合约是否为庞氏骗局合约。

[0004] 对于庞氏骗局合约检测,现有的技术方案包括:

[0005] 方案一:首先通过人工检查开源合约的代码,对庞氏骗局智能合约做出标记,根据庞氏骗局的实现形式将这类智能合约分为4个模式:基于数组的金字塔模式,基于树的金字塔模式,直接移交模式,瀑布模式。对于没有公开源代码的智能合约,通过比较合约字节码与庞氏骗局标记合约的字节码的相似度来进行判断。该方案的准确率低,方案一将智能合约字节码与标记过的庞氏骗局合约字节码进行相似度比较,由于受标记的庞氏骗局合约数量少,智能合约代码的复杂性高,该方法的检测准确率较低,难以检测多样化的庞氏骗局代码。

[0006] 方案二:通过提取智能合约的交易信息以及字节码特征,使用机器学习模型XBoost进行检测。对于合约的交易信息,该方法选取了7种简单交易信息,如合约余额,总交易数量,转入交易数量等。对于字节码信息,该方法使用了将字节码转换为操作码然后提取词频特征。最后,该方法将两种特征结合,使用XBoost算法训练机器学习模型,提供了一种庞氏骗局合约的自动化检测方法。该方案的检测率“虚高”,方案二的高检测率很大一部分依赖于对智能合约字节码特征的提取。然而实际上高检测率的背后是智能合约的高重复率:以太坊上有超过79.2%的智能合约都是重复的。因此使用基于合约代码的检测方法有很大一部分检测结果是来源于这些重复代码。导致这些方法表现出高检测率却难以应对新

的代码形式的庞氏骗局合约。

发明内容

[0007] 本发明提供了基于以太坊上交易时序信息的庞氏骗局智能合约检测方法，旨在解决上述问题。

[0008] 根据本申请实施例提供的基于以太坊上交易时序信息的庞氏骗局智能合约检测方法，包括以下步骤：

[0009] 步骤S1：对以太坊上的智能合约相关数据进行收集并标记庞氏骗局合约，得到智能合约及其标记的集合，以及涉及对应合约的交易集合；

[0010] 步骤S2：对于每个合约的交易集合，按照交易另一方的地址进行子集划分，根据智能合约与对应地址首笔交易的时间进行排序，得到排序后的交易子集；

[0011] 步骤S3：从划分后的交易集合中提取交易的时序特征，根据给定阈值进行交易记录的筛选；

[0012] 步骤S4：将时序特征和对应的合约标记时序序列预测模型LSTM网络的输入，得到经过训练的模型；

[0013] 步骤S5：使用训练模型对待检测的智能合约执行检测，得到是否是庞氏骗局合约的预测结果；

[0014] 其中，智能合约标记的集合为$\{(c_1,label_1),(c_2,label_2),(c_3,label_3),\cdots,(c_n,label_n)\}$，对应合约的交易集合为$\{T_1,T_2,T_3,\cdots,T_n\}$，其中$T_i$为第i个智能合约对应的交易集合。

[0015] 优选地，所述步骤S1包括：

[0016] 步骤S11：人工阅读合约源代码，根据合约是否实现庞氏骗局逻辑进行标记；

[0017] 步骤S12：通过交叉检查减少错误标记率，判断智能合约是否实现了庞氏骗局逻辑；

[0018] 步骤S13：若智能合约实现了庞氏骗局逻辑，则将此样本标记为正样本；

[0019] 步骤S14：若智能合约未实现庞氏骗局逻辑，则将此样本标记为负样本。

[0020] 优选地，所述步骤S2包括：

[0021] 步骤S21：对于每个智能合约，以与智能合约发生交易的账户地址为分类标准，将交易记录分为k个子集；

[0022] 步骤S22：按照每个账户与智能合约发生第一笔交易的时间，对k个子集进行排序，得到划分后的交易记录；

[0023] 步骤S23：将所有智能合约的交易记录集合起来，得到总的交易记录划分；

[0024] 其中，k为与智能合约发生交易的账户数量，划分后的交易记录为$\{T_{i1},T_{i2},\cdots,T_{ik}\}$，总的交易记录划分$\{T_{11},T_{12},\cdots,T_{n1},T_{n2},\cdots,T_{nk}\}$。

[0025] 优选地，所述步骤S3包括：从“使用后来投资者的投资作为前面投资者的利润”这一庞氏骗局的固有交易模式出发，对智能合约与每个交易账户的交易记录子集$\{T_{i1},T_{i2},\cdots,T_{ik}\}$，提取包括以下特征：

[0026] 转入交易的数量，交易源地址为对应交易账户，目的地址为该智能合约的交易数量；

[0027] 转出交易的数量,交易源地址为该智能合约,目的地址为对应交易账户的交易数量;

[0028] 智能合约与该账户总交易数量;

[0029] 第一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生第一笔交易的时间;

[0030] 最后一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生最后一笔交易的时间;

[0031] 发起第一笔交易时合约的余额;

[0032] 与该账户的第一笔交易是转入交易还是转出交易;

[0033] 交易总回报,转出交易的总金额减去转入交易的总金额;

[0034] 智能合约与该账户交易时间周期,最后一笔交易时间与第一笔交易时间的差值;

[0035] 对每个智能合约,通过上述步骤提取得到大小为(k,9)的时序特征矩阵X_i,对数据集中所有智能合约执行上述特征提取步骤,得到n个时序特征矩阵。

[0036] 优选地,所述步骤S4包括:

[0037] 步骤S41:判断合约发生交易的地址数量k与k_{min}、k_{max}的大小关系;

[0038] 步骤S42:如果与合约发生交易的地址数量k满足$k<k_{min}$,则认为合约没有足够的交易信息来帮助模型检测庞氏骗局合约,将这些合约排除在外;

[0039] 步骤S43:如果与合约发生交易的地址数量k满足$k_{min}<k<k_{max}$,则在大小为(k,9)的时序特征矩阵后补k_{max}-k行,9列的0元素作为填充,得到大小为(k_{max},9)的时序特征矩阵;

[0040] 步骤S44:如果与合约发生交易的地址数量k满足$k>k_{max}$,则从时序特征矩阵中均匀采样k_{max}行数据,得到大小为(k_{max},9)的时序特征矩阵;

[0041] 步骤S45:综合得到大小为(N,k_{max},9)的特征矩阵;

[0042] 在(N,k_{max},9)的特征矩阵中,N代表每个智能合约,k_{max}代表与该智能合约进行交易的账户,9代表代表该账户与该智能合约交易的特征。

[0043] 优选地,所述步骤S5包括:将提取到的智能合约交易时序特征与对应标记输入时序分类模型进行训练时,对LSTM网络进行参数配置和调整,在数据集上进行训练,测试以及验证。

[0044] 优选地,对于待分类的智能合约,所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法还包括:

[0045] 步骤S100:收集待分类智能合约的交易记录,对交易记录进行时序特征提取得到大小为(k,9)的时序特征矩阵;

[0046] 步骤S110:判断智能合约交易的账户数量k与k_{min}、k_{max}的大小关系;

[0047] 步骤S120:若与该智能合约交易的账户数量k满足$k<k_{max}$,则在时序特征矩阵后补0得到大小为(k_{max},9)的时序特征矩阵;

[0048] 步骤S130:若与该智能合约交易的账户数量k满足$k<k_{min}$,则提示预测结果可信度不高;

[0049] 步骤S140:若与该智能合约交易的账户数量k满足$k>k_{max}$,则对矩阵的行进行均匀抽样,得到大小为(k_{max},9)的时序特征矩阵;

[0050] 步骤S150:将得到的时序特征输入训练好的LSTM网络,得到预测结果。

[0051] 本申请实施例提供的技术方案可以包括以下有益效果:本申请设计了基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,本发明对智能合约交易的时序特征进行提取,用于庞氏骗局智能合约的检测,相比现有的几个庞氏骗局智能合约的检测方案,本发明使用了更加丰富的特征来表示智能合约的交易信息,从每个智能合约交易记录中提取了总共k*9个特征,本发明根据庞氏骗局的定义,从智能合约的交易数据出发提取交易信息时序特征,具有更好的扩展性。

[0052] 由于庞氏骗局的本质是“使用后来投资者的投资作为前面投资者的利润”这一交易模式,所以相比于从智能合约的字节码进行庞氏骗局的检测,本发明从合约的交易模式出发做检测可以更加容易理解该检测模型为什么有效,具有更好的可解释性。

[0053] 本发明中的庞氏骗局合约检测方法只基于智能合约交易信息时序特征,不依赖智能合约的代码信息。避免了由于以太坊上智能合约代码的高重复率带来的在小规模数据集上检测率“虚高”的问题。由于“使用后来投资者的投资作为前面投资者的利润”是庞氏骗局的本质特征,因此任何代码形式的庞氏骗局都应当符合一定的交易信息时序特征,因此本发明可以应对实际场景中代码形式不断更新的庞氏骗局合约,具有更好的健壮性和可扩展性。

附图说明

[0054] 为了更清楚地说明本发明实施例技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0055] 图1是本发明基于以太坊上交易时序信息的庞氏骗局智能合约检测方法的流程示意图;

[0056] 图2是本发明基于以太坊上交易时序信息的庞氏骗局智能合约检测方法中步骤S1的流程示意图;

[0057] 图3是本发明基于以太坊上交易时序信息的庞氏骗局智能合约检测方法中步骤S2的流程示意图;

[0058] 图4是本发明基于以太坊上交易时序信息的庞氏骗局智能合约检测方法中步骤S4的流程示意图;

[0059] 图5是本发明基于以太坊上交易时序信息的庞氏骗局智能合约检测方法中的另一流程示意图。

具体实施方式

[0060] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0061] 还应当理解,在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0062] 还应当进一步理解，在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

[0063] 请参阅图1，本发明公开了基于以太坊上交易时序信息的庞氏骗局智能合约检测方法10，所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法10包括以下步骤：

[0064] 步骤S1：对以太坊上的智能合约相关数据进行收集并标记庞氏骗局合约，得到智能合约及其标记的集合，以及涉及对应合约的交易集合；

[0065] 步骤S2：对于每个合约的交易集合，按照交易另一方的地址进行子集划分，根据智能合约与对应地址首笔交易的时间进行排序，得到排序后的交易子集；

[0066] 步骤S3：从划分后的交易集合中提取交易的时序特征，根据给定阈值进行交易记录的筛选；

[0067] 步骤S4：将时序特征和对应的合约标记时序序列预测模型LSTM网络的输入，得到经过训练的模型；

[0068] 步骤S5：使用训练模型对待检测的智能合约执行检测，得到是否是庞氏骗局合约的预测结果；

[0069] 其中，智能合约标记的集合为$\{(c_1,label_1),(c_2,label_2),(c_3,label_3),\cdots,(c_n,label_n)\}$，对应合约的交易集合为$\{T_1,T_2,T_3,\cdots,T_n\}$，其中$T_i$为第i个智能合约对应的交易集合，排序后的交易子集为$\{T_{11},T_{12},\cdots,T_{21},T_{22},\cdots,T_{nk}\}$，步骤S3中给定的阈值为$k_{min}$和$k_{max}$。

[0070] 所述的智能合约相关数据包括经验证的开源智能合约的源代码以及涉及对应智能合约的所有交易。智能合约源代码的收集可以通过在etherscan.io网站上爬取，智能合约的交易信息可以通过以太坊客户端Parity来收集。Parity除了能提供基础的区块数据之外，还提供了方便的API，直接获取某些难以从区块数据中获得的数据如内部交易等。

[0071] 所述步骤S4中，使用的LSTM网络的输入为大小$(k_{max},9)$的特征矩阵，输出为0或者1，分别代表智能合约是普通合约、庞氏骗局合约。将步骤S3中提取交易的时序特征和对应合约的标签，即合约是否是庞氏骗局合约，用于时序序列预测模型LSTM神经网络的训练过程，经过一系列训练可以得到庞氏骗局合约的分类模型。对于新的智能合约，将从合约中提取到的的特征矩阵作为模型的输入，模型可给出0或者1的输出，0对应合约为正常合约，1对应合约为庞氏骗局合约。

[0072] 采用此设计，本发明对智能合约交易的时序特征进行提取，用于庞氏骗局智能合约的检测，相比现有的几个庞氏骗局智能合约的检测方案，本发明使用了更加丰富的特征来表示智能合约的交易信息，从每个智能合约交易记录中提取了总共k*9个特征，本发明根据庞氏骗局的定义，从智能合约的交易数据出发提取交易信息时序特征，具有更好的扩展性。

[0073] 由于庞氏骗局的本质是“使用后来投资者的投资作为前面投资者的利润”这一交易模式，所以相比于从智能合约的字节码进行庞氏骗局的检测，本发明从合约的交易模式出发做检测可以更加容易理解该检测模型为什么有效，具有更好的可解释性。

[0074] 本发明中的庞氏骗局合约检测方法只基于智能合约交易信息时序特征，不依赖智能合约的代码信息。避免了由于以太坊上智能合约代码的高重复率带来的在小规模数据集上检测率“虚高”的问题。由于“使用后来投资者的投资作为前面投资者的利润”是庞氏骗局的本质特征，因此任何代码形式的庞氏骗局都应当符合一定的交易信息时序特征，因此本

发明可以应对实际场景中代码形式不断更新的庞氏骗局合约,具有更好的健壮性和可扩展性。

[0075] 请参阅图2,所述步骤S1包括:

[0076] 步骤S11:人工阅读合约源代码,根据合约是否实现庞氏骗局逻辑进行标记;

[0077] 步骤S12:通过交叉检查减少错误标记率,判断智能合约是否实现了庞氏骗局逻辑;

[0078] 步骤S13:若智能合约实现了庞氏骗局逻辑,则将此样本标记为正样本;

[0079] 步骤S14:若智能合约未实现庞氏骗局逻辑,则将此样本标记为负样本。

[0080] 请参阅图3,所述步骤S2包括:

[0081] 步骤S21:对于每个智能合约,以与智能合约发生交易的账户地址为分类标准,将交易记录分为k个子集;

[0082] 步骤S22:按照每个账户与智能合约发生第一笔交易的时间,对k个子集进行排序,得到划分后的交易记录;

[0083] 步骤S23:将所有智能合约的交易记录集合起来,得到总的交易记录划分;

[0084] 其中,k为与智能合约发生交易的账户数量,划分后的交易记录为$\{T_{i1},T_{i2},\cdots,T_{ik}\}$,总的交易记录划分$\{T_{11},T_{12},\cdots,T_{n1},T_{n2},\cdots,T_{nk}\}$。

[0085] 其中,所述步骤S3包括:从“使用后来投资者的投资作为前面投资者的利润”这一庞氏骗局的固有交易模式出发,对智能合约与每个交易账户的交易记录子集$\{T_{i1},T_{i2},\cdots,T_{ik}\}$,提取包括以下特征:

[0086] 转入交易的数量,交易源地址为对应交易账户,目的地址为该智能合约的交易数量;

[0087] 转出交易的数量,交易源地址为该智能合约,目的地址为对应交易账户的交易数量;

[0088] 智能合约与该账户总交易数量;

[0089] 第一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生第一笔交易的时间;

[0090] 最后一笔交易时间,以智能合约创建时间为基准,智能合约与对应交易账户发生最后一笔交易的时间;

[0091] 发起第一笔交易时合约的余额;

[0092] 与该账户的第一笔交易是转入交易还是转出交易;

[0093] 交易总回报,转出交易的总金额减去转入交易的总金额;

[0094] 智能合约与该账户交易时间周期,最后一笔交易时间与第一笔交易时间的差值;

[0095] 对每个智能合约,通过上述步骤提取得到大小为(k,9)的时序特征矩阵X_i,对数据集中所有智能合约执行上述特征提取步骤,得到n个时序特征矩阵。

[0096] 请参阅图4,所述步骤S4包括:

[0097] 步骤S41:判断合约发生交易的地址数量k与k_{min}、k_{max}的大小关系;

[0098] 步骤S42:如果与合约发生交易的地址数量k满足$k<k_{min}$,则认为合约没有足够的交易信息来帮助模型检测庞氏骗局合约,将这些合约排除在外;

[0099] 步骤S43:如果与合约发生交易的地址数量k满足$k_{min}<k<k_{max}$,则在大小为(k,9)的

时序特征矩阵后补k_{max}-k行,9列的0元素作为填充,得到大小为(k_{max},9)的时序特征矩阵;

[0100] 步骤S44:如果与合约发生交易的地址数量k满足k>k_{max},则从时序特征矩阵中均匀采样k_{max}行数据,得到大小为(k_{max},9)的时序特征矩阵;

[0101] 步骤S45:综合得到大小为(N,k_{max},9)的特征矩阵;

[0102] 在(N,k_{max},9)的特征矩阵中,N代表每个智能合约,k_{max}代表与该智能合约进行交易的账户,9代表代表该账户与该智能合约交易的特征。所述步骤S4中使用的LSTM神经网络以大小为(k_{max},9)的矩阵作为输入。由于步骤S3中,从每个智能合约提取的特征矩阵大小取决于智能合约本身的交易记录,每个智能合约的特征矩阵大小不同,因此需要对特征矩阵大小进行标准化。要将得到的大小为(k,9)的矩阵转换为大小为(k_{max},9)的标准大小矩阵,需要对特征矩阵进行数据的增添或删减,最终得到大小为(k_{max},9)的矩阵。

[0103] 其中,所述步骤S5包括:将提取到的智能合约交易时序特征与对应标记输入时序分类模型进行训练时,对LSTM网络进行参数配置和调整,在数据集上进行训练,测试以及验证。

[0104] 请参阅图5,对于待分类的智能合约,所述基于以太坊上交易时序信息的庞氏骗局智能合约检测方法10还包括:

[0105] 步骤S100:收集待分类智能合约的交易记录,对交易记录进行时序特征提取得到大小为(k,9)的时序特征矩阵;

[0106] 步骤S110:判断智能合约交易的账户数量k与k_{min}、k_{max}的大小关系;

[0107] 步骤S120:若与该智能合约交易的账户数量k满足k<k_{max},则在时序特征矩阵后补0得到大小为(k_{max},9)的时序特征矩阵;

[0108] 步骤S130:若与该智能合约交易的账户数量k满足k<k_{min},则提示预测结果可信度不高;

[0109] 步骤S140:若与该智能合约交易的账户数量k满足k>k_{max},则对矩阵的行进行均匀抽样,得到大小为(k_{max},9)的时序特征矩阵;

[0110] 步骤S150:将得到的时序特征输入训练好的LSTM网络,得到预测结果。

[0111] 本申请实施例提供的技术方案可以包括以下有益效果:本申请设计了基于以太坊上交易时序信息的庞氏骗局智能合约检测方法,本发明对智能合约交易的时序特征进行提取,用于庞氏骗局智能合约的检测,相比现有的几个庞氏骗局智能合约的检测方案,本发明使用了更加丰富的特征来表示智能合约的交易信息,从每个智能合约交易记录中提取了总共k*9个特征,本发明根据庞氏骗局的定义,从智能合约的交易数据出发提取交易信息时序特征,具有更好的扩展性。

[0112] 由于庞氏骗局的本质是“使用后来投资者的投资作为前面投资者的利润”这一交易模式,所以相比于从智能合约的字节码进行庞氏骗局的检测,本发明从合约的交易模式出发做检测可以更加容易理解该检测模型为什么有效,具有更好的可解释性。

[0113] 本发明中的庞氏骗局合约检测方法只基于智能合约交易信息时序特征,不依赖智能合约的代码信息。避免了由于以太坊上智能合约代码的高重复率带来的在小规模数据集上检测率“虚高”的问题。由于“使用后来投资者的投资作为前面投资者的利润”是庞氏骗局的本质特征,因此任何代码形式的庞氏骗局都应当符合一定的交易信息时序特征,因此本发明可以应对实际场景中代码形式不断更新的庞氏骗局合约,具有更好的健壮性和可扩展

性。

[0114] 以上所述，仅为本发明的具体实施方式，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揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改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以权利要求的保护范围为准。

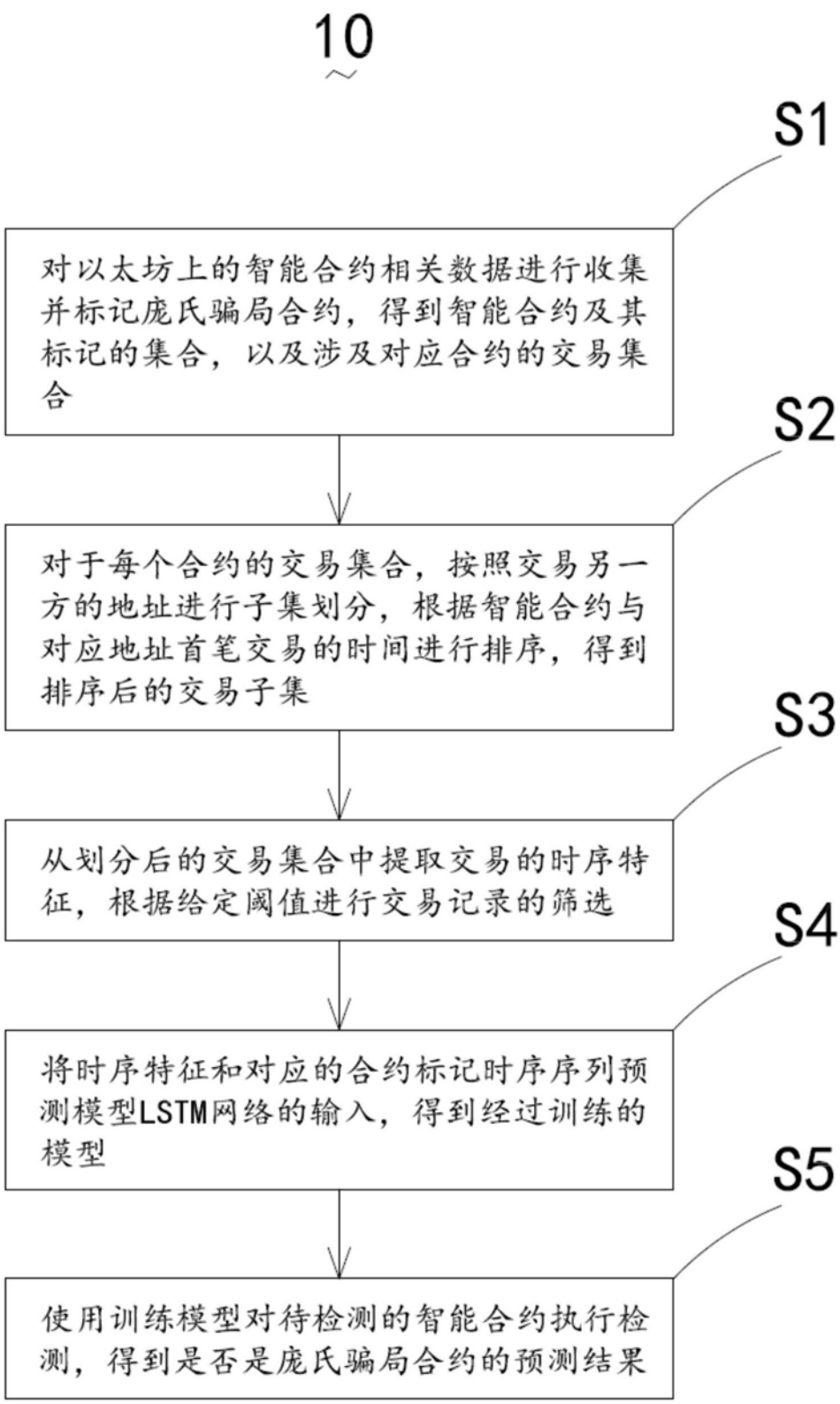

10
S1
对以太坊上的智能合约相关数据进行收集并标记庞氏骗局合约，得到智能合约及其标记的集合，以及涉及对应合约的交易集合
S2
对于每个合约的交易集合，按照交易另一方的地址进行子集划分，根据智能合约与对应地址首笔交易的时间进行排序，得到排序后的交易子集
S3
从划分后的交易集合中提取交易的时序特征，根据给定阈值进行交易记录的筛选
S4
将时序特征和对应的合约标记时序序列预测模型LSTM网络的输入，得到经过训练的模型
S5
使用训练模型对待检测的智能合约执行检测，得到是否是庞氏骗局合约的预测结果

图1

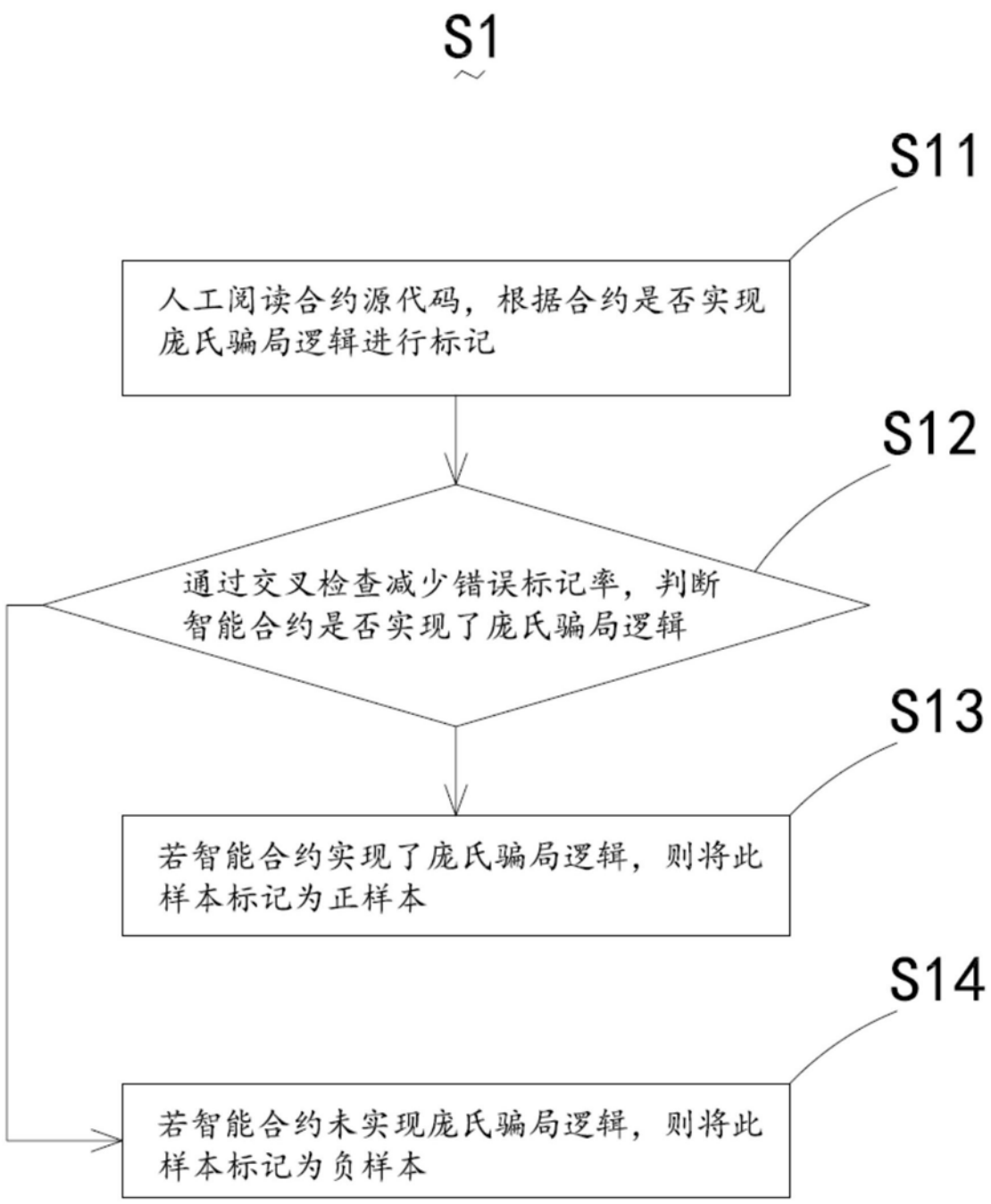

S1
S11
人工阅读合约源代码，根据合约是否实现庞氏骗局逻辑进行标记
S12
通过交叉检查减少错误标记率，判断智能合约是否实现了庞氏骗局逻辑
S13
若智能合约实现了庞氏骗局逻辑，则将此样本标记为正样本
S14
若智能合约未实现庞氏骗局逻辑，则将此样本标记为负样本

图2

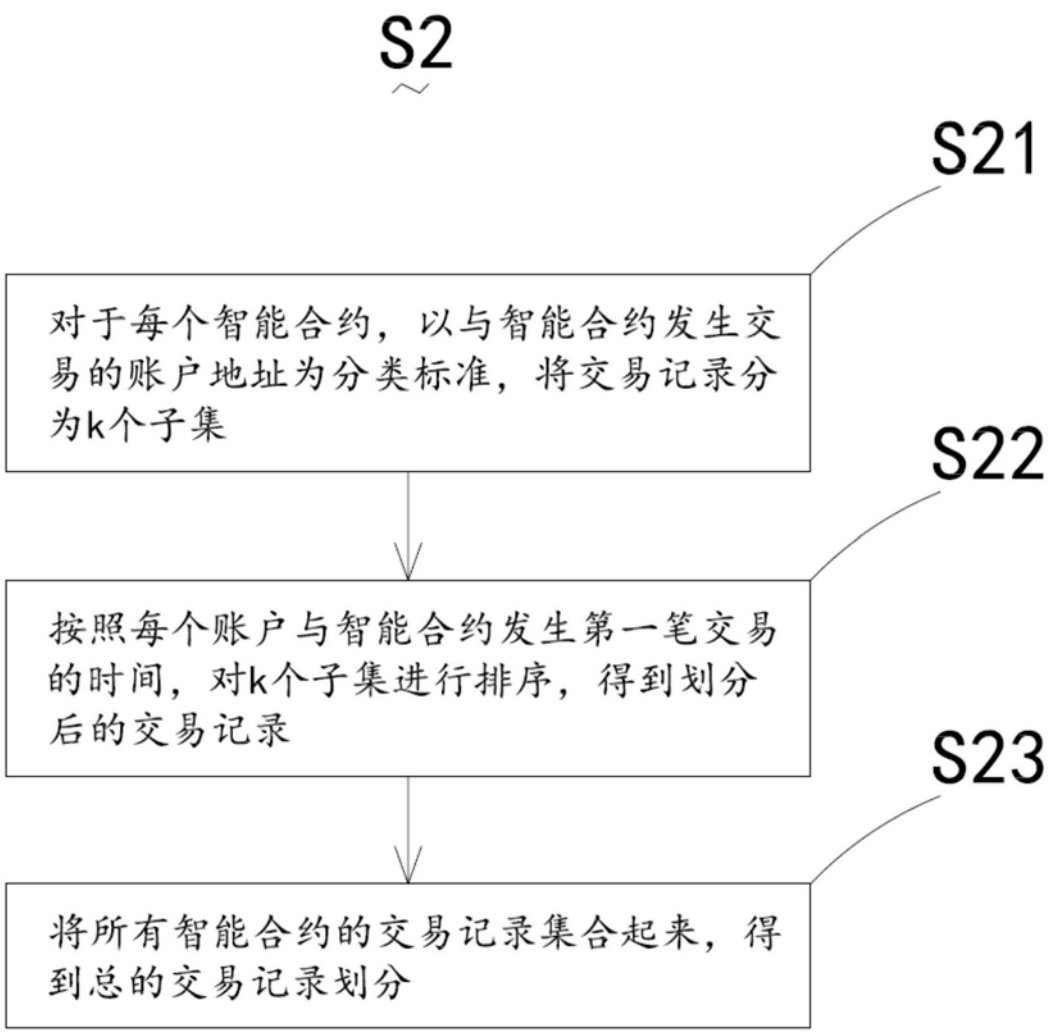

S2
S21
对于每个智能合约，以与智能合约发生交易的账户地址为分类标准，将交易记录分为k个子集
S22
按照每个账户与智能合约发生第一笔交易的时间，对k个子集进行排序，得到划分后的交易记录
S23
将所有智能合约的交易记录集合起来，得到总的交易记录划分

图3

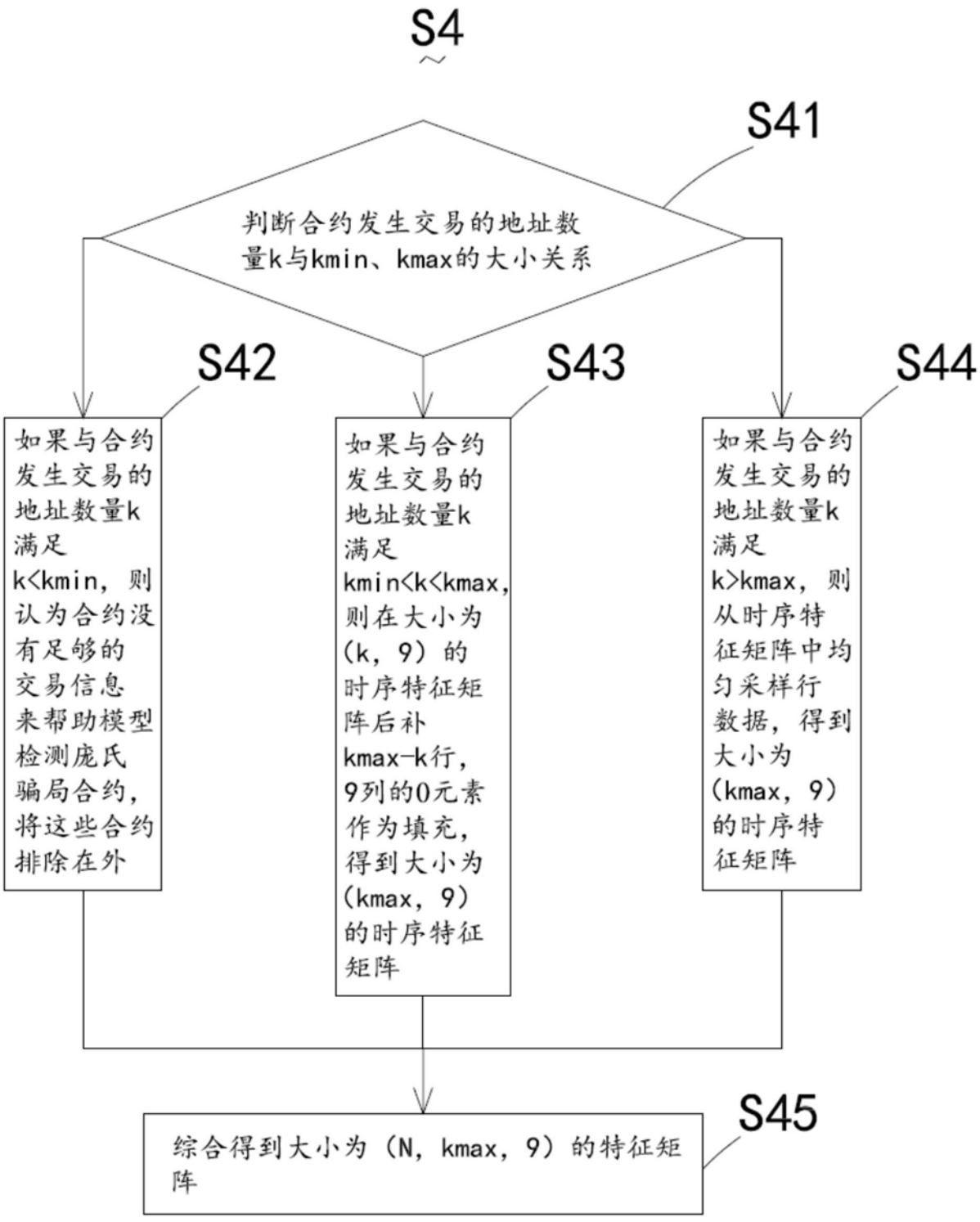

S4
S41
判断合约发生交易的地址数量k与kmin、kmax的大小关系
S42
如果与合约发生交易的地址数量k满足k<kmin，则认为合约没有足够的交易信息来帮助模型检测庞氏骗局合约，将这些合约排除在外
S43
如果与合约发生交易的地址数量k满足kmin<k<kmax，则在大小为（k，9）的时序特征矩阵后补kmax-k行，9列的0元素作为填充，得到大小为（kmax，9）的时序特征矩阵
S44
如果与合约发生交易的地址数量k满足k>kmax，则从时序特征矩阵中均匀采样行数据，得到大小为（kmax，9）的时序特征矩阵
S45
综合得到大小为（N，kmax，9）的特征矩阵

图4

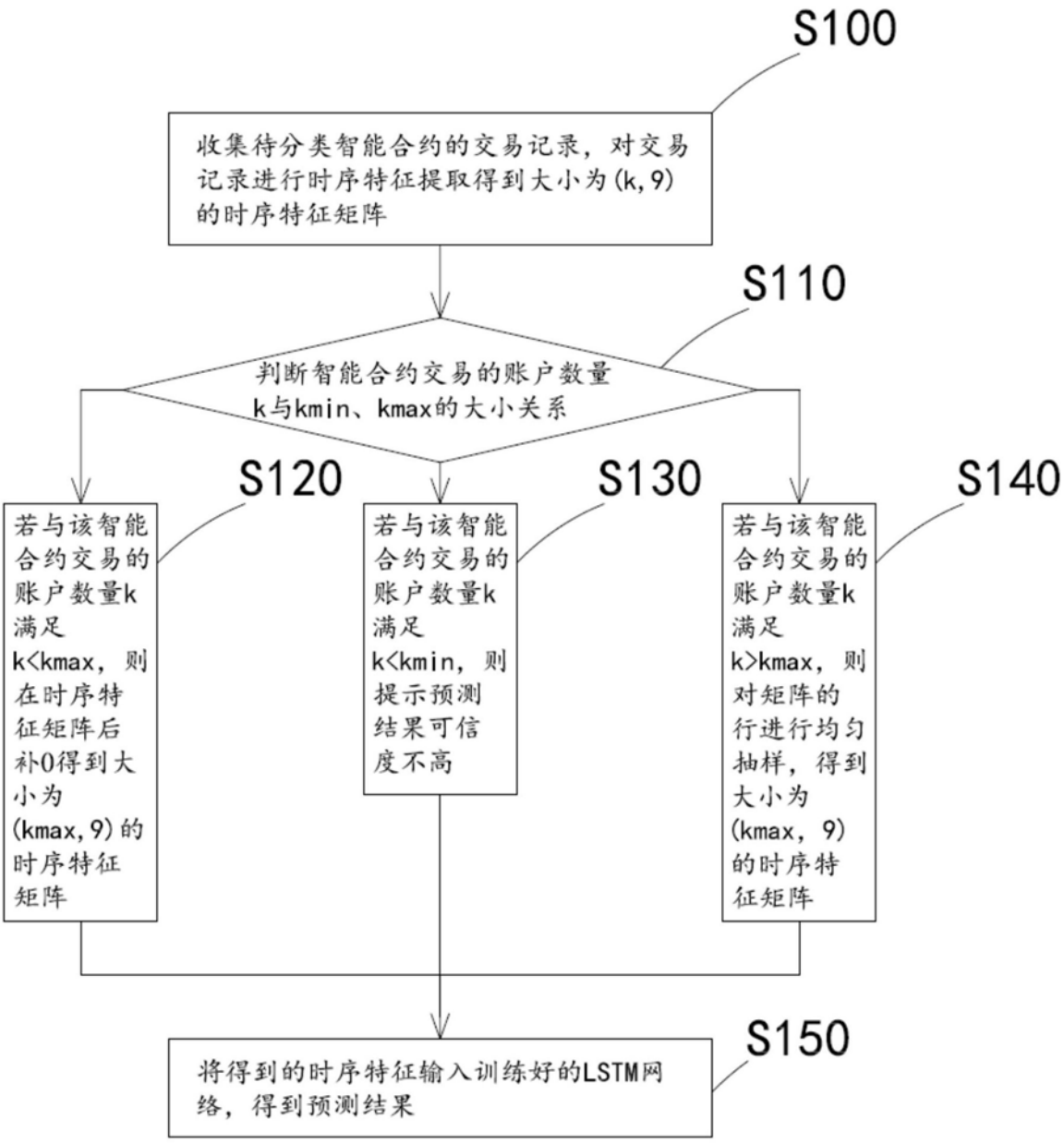

S100
收集待分类智能合约的交易记录，对交易记录进行时序特征提取得到大小为(k,9)的时序特征矩阵
S110
判断智能合约交易的账户数量k与kmin、kmax的大小关系
S120
若与该智能合约交易的账户数量k满足k<kmax，则在时序特征矩阵后补0得到大小为(kmax,9)的时序特征矩阵
S130
若与该智能合约交易的账户数量k满足k<kmin，则提示预测结果可信度不高
S140
若与该智能合约交易的账户数量k满足k>kmax，则对矩阵的行进行均匀抽样，得到大小为(kmax，9)的时序特征矩阵
S150
将得到的时序特征输入训练好的LSTM网络，得到预测结果

图5