(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 112699964 A (43) 申请公布日 2021. 04. 23

(21)申请号 202110042468.7

(22) 申请日 2021.01.13

(71) 申请人 成都链安科技有限公司 地址 610000 四川省成都市中国(四川)自 由贸易试验区成都高新区世纪城南路 599号7栋5层505A、505B号

(72) 发明人 杨霞 郭文生

(74) 专利代理机构 成都云纵知识产权代理事务 所(普通合伙) 51316

代理人 熊曦 刘沙粒

(51) Int.CI.

GO6K 9/62 (2006.01)

G06Q 20/06 (2012.01)

G06Q 20/40 (2012.01)

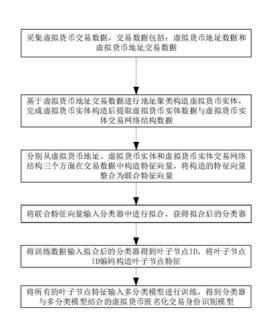
权利要求书2页 说明书9页 附图2页

(54) 发明名称

模型构建方法、系统、装置、介质、交易身份识别方法

(57) 摘要

本发明公开了模型构建方法、系统、装置、介质、交易身份识别方法,涉及区块链虚拟货币领域,包括:采集虚拟货币交易数据;聚类构造虚拟货币实体,提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量;将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型,本发明能够解决虚拟货币地址的匿名性问题,自动识别不同虚拟货币地址的身份。



CN 112699964 A

1.模型构建方法,其特征在于,所述方法包括:

采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据;

基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;

分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量;

将联合特征向量输入分类器中进行拟合,获得拟合后的分类器:

将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征:

将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。

- 2.根据权利要求1所述的模型构建方法,其特征在于,本方法从虚拟货币交易系统中采集交易数据。
- 3.根据权利要求1所述的模型构建方法,其特征在于,本方法通过独热编码器对叶子节点的ID进行独热编码构造出叶子节点特征。
- 4.根据权利要求1所述的模型构建方法,其特征在于,本方法中的多分类模型为Softmax多分类模型。
- 5.根据权利要求1所述的模型构建方法,其特征在于,本方法采用数据统计性描述方式构造特征向量。
 - 6.根据权利要求1所述的模型构建方法,其特征在于,本方法中的分类器为随机森林。
 - 7.模型构建系统,其特征在于,所述系统包括:

采集单元,用于采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据;

聚类单元,用于基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;

联合特征向量构造单元,用于分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量;

拟合单元,用于将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;

叶子节点特征构造单元,用于将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;

训练单元,用于将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。

8.虚拟货币匿名化交易身份识别方法,其特征在于,所述方法包括:

获得待处理虚拟货币交易数据:

将待处理虚拟货币交易数据输入权利要求1-6中任意一个所述模型构建方法构建的虚拟货币匿名化交易身份识别模型中;

虚拟货币匿名化交易身份识别模型输出虚拟货币匿名化交易身份识别结果。

9.一种模型构建装置,包括存储器、处理器以及存储在所述存储器中并可在所述处理

器上运行的计算机程序,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 1-6中任意一个所述模型构建方法的步骤。

10.一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-6中任意一个所述模型构建方法的步骤。

模型构建方法、系统、装置、介质、交易身份识别方法

技术领域

[0001] 本发明涉及区块链虚拟货币领域,具体地,涉及模型构建方法、系统、装置、介质、交易身份识别方法。

背景技术

[0002] 由于虚拟货币的交易具有匿名化的特点,有不法分子利用这一特点进行一些违法活动,如洗钱等,导致相关部门无法监管,为了便于相关部门对虚拟货币的交易进行监管则需要对虚拟货币的交易进行去匿名化处理。

[0003] 现有技术中有采用机器学习的方式进行虚拟货币的交易去匿名化处理,传统的机器学习可以轻松处理上亿条数据,但是学习能力十分有限,需要大量的特征工程来增加模型的学习能力。但大量的特征工程耗时耗力同时并不一定会带来效果提升。因此,如何自动发现有效的特征、特征组合,弥补人工经验不足,缩短传统机器学习特征实验周期,是亟需解决的问题。

发明内容

[0004] 为解决上述问题,本发明提供了模型构建方法、系统、装置、介质、交易身份识别方法,本发明能够解决虚拟货币地址的匿名性问题,自动识别不同虚拟货币地址的身份,从而提高安全执法部门的工作效率。

[0005] 为实现上述目的,本发明提供了模型构建方法,所述方法包括:

[0006] 采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据:

[0007] 基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;

[0008] 分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量;

[0009] 将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;

[0010] 将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;其中,本方法中原始特征通过随机森林模型转换后,能够自动的挖掘一些重要特征以及重要特征组合,树模型上到叶子节点的每一条路径就相当于一种特征的组合,因此随机森林模型能够自动的做特征的非线性变换和特征组合,减少了大量的人为特征工程,主要作用是提升特征的准确性,筛选有效特征;

[0011] 将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。由于虚拟货币地址身份识别的场景属于多分类场景,此步骤的作用主要是用来对经过随机森林进行原始特征的非线性变换与特征组合的新特征进行比特币交易地址的多分类识别,锻炼混合算法的使用,避免单一算法的局限性。

[0012] 优选的,本方法从虚拟货币交易系统中采集交易数据。

[0013] 优选的,本方法通过独热编码器对叶子节点的ID进行独热编码构造出叶子节点特征。原始数据经过随机森林转换过后的数据为每棵树的叶子节点ID(每棵树一个对应的数字),不能直接输入Softmax中进行训练,因此需要把叶子节点的数据向量化,利用热独编码可以将叶子节点的数据向量化,利用热度编码构造softmax的特征数据。

[0014] 优选的,本方法中的多分类模型为Softmax多分类模型。

[0015] 优选的,本方法采用数据统计性描述方式构造特征向量。

[0016] 优选的,本方法中的分类器为随机森林。

[0017] 本发明还提供了模型构建系统,所述系统包括:

[0018] 采集单元,用于采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据;

[0019] 聚类单元,用于基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;

[0020] 联合特征向量构造单元,用于分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量:

[0021] 拟合单元,用于将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;

[0022] 叶子节点特征构造单元,用于将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;

[0023] 训练单元,用于将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。

[0024] 本发明还提供了虚拟货币匿名化交易身份识别方法,所述方法包括:

[0025] 获得待处理虚拟货币交易数据;

[0026] 将待处理虚拟货币交易数据输入所述模型构建方法构建的虚拟货币匿名化交易身份识别模型中:

[0027] 虚拟货币匿名化交易身份识别模型输出虚拟货币匿名化交易身份识别结果。

[0028] 本发明还提供了一种模型构建装置,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现所述模型构建方法的步骤。

[0029] 本发明还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现所述模型构建方法的步骤。

[0030] 本发明提供的一个或多个技术方案,至少具有如下技术效果或优点:

[0031] 本发明通过随机森林生成的特征,可直接作为Softmax逻辑回归模型的特征使用,省去人工处理分析特征的环节,LR的输入特征完全依赖于通过随机森林得到的特征,因此在特征工程中,依靠随机森林进行自动特征工程的工作,且随机森林算法的特点正好可以用来发掘有区分度的特征、特征组合,减少特征工程中人力成本。

附图说明

[0032] 此处所说明的附图用来提供对本发明实施例的进一步理解,构成本发明的一部分,并不构成对本发明实施例的限定;

[0033] 图1为模型构建方法的流程示意图:

[0034] 图2为比特币交易示意图:

[0035] 图3为随机森林+Softmax模型流程示意图:

[0036] 图4为模型构建系统的组成示意图。

具体实施方式

[0037] 为了能够更清楚地理解本发明的上述目的、特征和优点,下面结合附图和具体实施方式对本发明进行进一步的详细描述。需要说明的是,在相互不冲突的情况下,本发明的实施例及实施例中的特征可以相互组合。

[0038] 在下面的描述中阐述了很多具体细节以便于充分理解本发明,但是,本发明还可以采用其他不同于在此描述范围内的其他方式来实施,因此,本发明的保护范围并不受下面公开的具体实施例的限制。

[0039] 本领域技术人员应理解的是,在本发明的揭露中,术语"纵向"、"横向"、"上"、"下"、"前"、"后"、"左"、"右"、"竖直"、"水平"、"顶"、"底""内"、"外"等指示的方位或位置关系, 其仅是为了便于描述本发明和简化描述, 而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作, 因此上述术语不能理解为对本发明的限制。

[0040] 可以理解的是,术语"一"应理解为"至少一"或"一个或多个",即在一个实施例中,一个元件的数量可以为一个,而在另外的实施例中,该元件的数量可以为多个,术语"一"不能理解为对数量的限制。

[0041] 实施例一

[0042] 请参考图1,图1为模型构建方法的流程示意图,本发明实施例一提供了模型构建方法,所述方法包括:

[0043] 采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据;

[0044] 基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据:

[0045] 分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量;

[0046] 将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;

[0047] 将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;

[0048] 将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。

[0049] 其中,本发明实施例中虚拟货币为比特币。

[0050] 其中,本发明实施例中,本方法从虚拟货币交易系统中采集交易数据,在具体实施过程中也可任意从其他途径获得交易数据,本发明对交易数据的获取方式、途径、手段不进行具体的限定。

[0051] 其中,本发明实施例中,本方法通过独热编码器对叶子节点的ID进行独热编码构

造出叶子节点特征。其中,在实际应用中也可以采用其他编码器或编码方式进行编码,本发明对具体的编码方式不进行具体的限定。

[0052] 其中,本发明实施例中,本方法中的多分类模型为Softmax多分类模型。在实际应用中也可以为其他类型的多分类模型,本发明对分类模型的具体类型不进行限定。

[0053] 其中,本发明实施例中,本方法中的分类器为随机森林。在实际应用中也可以为其他类型的分类器,本发明对分类器的具体类型不进行限定。

[0054] 下面以虚拟货币为比特币举例对本发明进行具体介绍:

[0055] 图2为比特币交易示意图,数据直接来源于区块链全节点,并展示了比特币的流向。图2中的各个顶点代表了比特币链上的比特币地址 $(\alpha_1,\alpha_2,\alpha_3,\ldots,\alpha_N)$ 与相关交易 $(tx_1,tx_2,tx_3,\ldots,tx_N)$,无数这样的交易形成庞大的交易网络拓扑图。在比特币交易系统中,一个用户会使用单个或多个地址,在本发明中申请人引入比特币交易实体的概念,实体 (E_1,E_2,E_3,\ldots,E_N) 指的在逻辑上属于同一个用户的地址集合,也可理解成一个实体代表一个用户。

[0056] 本方法包括特征抽取过程和模型构建过程。

[0057] 特征抽取过程:在比特币交易系统中,当用户支付额度超过了用户钱包中每一个可用地址中比特币的数量时,为了避免执行多笔交易完成支付造成交易费用方面的损失,用户会从钱包中选择多个比特币地址聚合在一起进行匹配支付,实现多输入交易,因此这笔多输入地址的交易中的所有输入地址属于同一个实体(用户)。因此在构建比特币匿名化交易身份识别方法的时候,首先对地址进行实体聚类,然后分别从比特币地址、比特币实体、比特币实体交易网络结构(motif)的三个方面在海量的交易数据中去构造特征并整合成联合特征向量。

[0058] 模型构建过程:首先将联合特征向量输入随机森林中进行拟合,然后就可以得到多棵已经非常了解样本数据的树。然后将训练数据通过随机森林,这样每个样本数据输入树模型中就会得到叶子节点的ID,通过OneHotEncoder(独热编码器)对叶子节点的ID进行onehot(独热)编码就可以构造出叶子节点特征。最后将所有的叶子节点特征输入Softmax多分类模型进行训练,最终得到随机森林与Softmax结合的模型。

[0059] 实施例二

[0060] 请参考图3,图3为随机森林+Softmax模型流程示意图。本方法包括以下步骤:

[0061] 步骤1:特征工程

[0062] 本发明提出实体的概念,实体指的在逻辑上属于同一个用户的地址集合,也可理解成一个实体代表一个用户。首先把比特币地址聚合成实体,然后分别根据比特币地址、实体、实体交易网络结构进行特征抽取,并合并成联合向量,具体的特征抽取手段为数据统计性描述,抽取更能够解释业务场景的字段,详细的特征数据如表1-表3所示。

[0063] 表1地址特征

	特征名称	特征解释				
	余额	地址当前未花费的比特币				
	转入金额	地址接收的比特币总额				
	转出金额	地址转出的比特币总额				
	平均每笔转入金额	平均每笔交易中地址接收的比				
	十均母毛特八金领	特币				
	亚特复数共山入鄉	平均每笔交易中地址转出的比				
[0064]	平均每笔转出金额	特币				
	转入交易数量	地址接收比特币的交易数量				
	转出交易数量	地址转出比特币的交易数量				
	入度	地址转入交易中的交易地址对				
	八及	手				
	出度	地址转出交易中的交易地址对				
		手				

[0045]	主0分仔灶红					
[0065]	表2实体特征					
	特征名称	特征解释				
	余额	实体对应地址的比特币余额总				
[0066]		和				
	转入金额	实体对应地址转出的比特币总				
		和				
	转出金额	实体对应地址转出的比特币总				
		和				
	可以与标件工人的	平均每笔交易中实体接收的比				
	平均每笔转入金额	特币				
	Till to the left of the Alexander	平均每笔交易中实体转出的比				
[00/7]	平均每笔转出金额	特币				
	转入交易数量	实体接收比特币的交易数量				
	转出交易数量	实体转出比特币的交易数量				
	口护铁头大目眩牙	实体每天接收比特币的交易数				
	日均转入交易频次	量				
	日均转出交易频次	实体每天转出比特币的交易数				
[0067]		量				
	入度	实体转入交易中的交易实体对				
		手				
	出度	实体转出交易中的交易实体对				
		手				
	PageRank	实体交易网络中实体重要性				
		实体对交易网络资金流影响程				
	介数中心性	度				
	实体寿命	实体交易起止时间长度				
		100				

[0068] 表3网络结构特征

特征名称

[0069]

不同 motif 结构的记录条数 不同 motif 结构的总交易费用 不同 motif 结构的平均交易费用 不同 motif 结构的平均输入的地址数量 不同 motif 结构的平均输出的地址数量 不同 motif 结构的平均输出的地址数量

[0070] 步骤2:随机森林

[0071] 将联合特征向量输入随机森林中进行拟合,可以得到多棵已经非常了解样本数据的树。然后将训练数据通过随机森林,这样每个样本数据输入树模型中就会得到叶子节点的ID,通过热度编码(OneHotEncoder)对叶子节点的ID进行OneHotEncoder可以构造出叶子节点特征数据;原始特征通过随机森林模型转换后,能够自动的挖掘一些重要特征以及重要特征组合,树模型上到叶子节点的每一条路径就相当于一种特征的组合,因此随机森林模型能够自动的做特征的非线性变换和特征组合,减少了大量的人为特征工程,主要作用是提升特征的准确性,筛选有效特征。而原始数据经过随机森林转换过后的数据为每棵树的叶子节点ID(每棵树一个对应的数字),不能直接输入Softmax中进行训练,因此需要把叶子节点的数据向量化,而热独编码为常用的向量化方法,因此热度编码被选来构造softmax的特征数据。

[0072] 步骤3:Softmax

[0073] 将随机森林拟合出来的叶子节点特征数据输入Softmax模型进行训练得到比特币地址的多分类模型,从而达到比特币地址的身份识别。由于比特币地址身份识别的场景属于多分类场景,此步骤的作用主要是用来对经过随机森林进行原始特征的非线性变换与特征组合的新特征进行比特币交易地址的多分类识别,锻炼混合算法的使用,避免单一算法的局限性。

[0074] 本发明实施例与现有技术的区别在于提出了将实体概念对比特币地址进行聚类; 并且为了提取到更全面,更有效的特征数据,本发明实施例提出联合特征的提取方法,构造 联合特征向量;同时本发明实施例引用了随机森林与Softmax模型相结合,从而达到通过模 型自动发掘有区分度的特征、特征组合,减少特征工程中人力成本,其次锻炼混合算法的使 用,避免单一算法的局限性。

[0075] 为了表面本发明的效果,申请人进行了相应的试验对比,对比结果如表4所示。

[0076] 表四:实验方案的明细结果

			实体特征分类模型			联合特征分类模型		
	Class	Model	P	R	F1-score	P	R	F1-score
	交易所	RF	0.67	0.88	0.76	0.96	0.94	0.95
[0077]	服务商	RF	0.67	0.11	0.18	0.89	0.75	0.81
	赌博	RF	0.56	0.78	0.65	0.95	0.88	0.91
	暗网	RF	0.98	0.82	0.89	1.00	0.92	0.96
	混币	RF	1.00	1.00	1.00	1.00	1.00	1.00
	矿池	RF	0.60	0.43	0.50	1.00	0.75	0.86
	交易所	RF+Softmax	0.68	0.94	0.79	0.97	0.96	0.96
	服务商	RF+Softmax	0.50	0.19	0.28	0.92	0.73	0.81
	赌博	RF+Softmax	0.65	0.68	0.67	0.98	0.92	0.95
	暗网	RF+Softmax	0.98	0.89	0.93	1.00	0.95	0.97
	混币	RF+Softmax	1.00	1.00	1.00	1.00	1.00	1.00
	矿池	RF+Softmax	1.00	0.50	0.67	1.00	0.87	0.93

[0078] 为了比较本发明提出的方法的优点与缺点,本发明实施例进行两个不同的实验。在第一个实验中,本发明建立一个比较简单的分类器,称为实体特征分类器,在这个分类器中,模型训练数据特征只包括从海量实体交易数据中提取的实体特征。在第二次实验中,本发明基于实体特征分类器,构建一个复杂的分类器,称为联合特征分类器,流程如图3所示,主要结合从海量比特币交易数据中提取出来的地址方面的特征、实体方面的特征与网络结构方面的特征作为模型的输入,然后输入不到不同的算法中进行训练。实验结果表4.

[0079] 1.通过实验对比查准率 (precision)、召回率 (recall) 与F1分数 (F1_score) 三个评价指标,随机森林 (RF) 与Softmax逻辑回归结合的方案优于单独使用随机森林 (RF),表明采用随机森林 (RF) 的进行特征工程抽取出有效特征的技术手段,再结合多分类模型Softmax逻辑回归进行训练,能够大幅提升分类器的准确率,避免使用单一算法的局限性。

[0080] 2.单一的使用实体特征的模型分类效果,远远低于使用联合向量的作为特征的模型,因此可表明采用联合特征构造方案的技术手段,能够更全面,更有效的描述比特币地址的交易行为

[0081] 实施例三

[0082] 请参考图4,图4为模型构建系统的组成示意图,本发明实施例三提供了模型构建系统,所述系统包括:

[0083] 采集单元,用于采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和虚拟货币地址交易数据::

[0084] 聚类单元,用于基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体,完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实体交易网络结构数据;

[0085] 联合特征向量构造单元,用于分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络结构三个方面在交易数据中构造特征向量,将构造的特征向量整合为联合特征向量:

[0086] 拟合单元,用于将联合特征向量输入分类器中进行拟合,获得拟合后的分类器;

[0087] 叶子节点特征构造单元,用于将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点ID编码构造叶子节点特征;

[0088] 训练单元,用于将所有的叶子节点特征输入多分类模型进行训练,得到分类器与多分类模型结合的虚拟货币匿名化交易身份识别模型。

[0089] 实施例四

[0090] 本发明实施例四提供了虚拟货币匿名化交易身份识别方法,所述方法包括:

[0091] 获得待处理虚拟货币交易数据:

[0092] 将待处理虚拟货币交易数据输入所述模型构建方法构建的虚拟货币匿名化交易身份识别模型中;

[0093] 虚拟货币匿名化交易身份识别模型输出虚拟货币匿名化交易身份识别结果。

[0094] 实施例五

[0095] 本发明实施例五还提供了一种模型构建装置,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现所述模型构建方法的步骤。

[0096] 其中,所述处理器可以是中央处理器(CPU,Central Processing Unit),还可以是其他通用处理器、数字信号处理器(digital signal processor)、专用集成电路(Application Specific Integrated Circuit)、现成可编程门阵列(Fieldprogrammable gate array)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0097] 所述存储器可用于存储所述计算机程序和/或模块,所述处理器通过运行或执行存储在所述存储器内的数据,实现发明中模型构建装置的各种功能。所述存储器可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等。此外,存储器可以包括高速随机存取存储器、还可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡,安全数字卡,闪存卡、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0098] 实施例六

[0099] 本发明实施例六还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现所述模型构建方法的步骤。

[0100] 所述模型构建装置如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机程序可存储于一计算机可读存介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码、对象代码形式、可执行文件或某些中间形式等。所述计算机可读取介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器、随机存储器、点载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减。

[0101] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0102] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精

神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

采集虚拟货币交易数据,交易数据包括:虚拟货币地址数据和 虚拟货币地址交易数据

基于虚拟货币地址交易数据进行地址聚类构造虚拟货币实体, 完成虚拟货币实体构造后提取虚拟货币实体数据与虚拟货币实 体交易网络结构数据

分别从虚拟货币地址、虚拟货币实体和虚拟货币实体交易网络 结构三个方面在交易数据中构造特征向量,将构造的特征向量 整合为联合特征向量

将联合特征向量输入分类器中进行拟合,获得拟合后的分类器

将训练数据输入拟合后的分类器得到叶子节点ID,将叶子节点 ID编码构造叶子节点特征

将所有的叶子节点特征输入多分类模型进行训练,得到分类器 与多分类模型结合的虚拟货币匿名化交易身份识别模型

图1

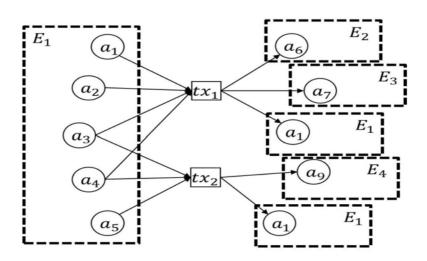


图2

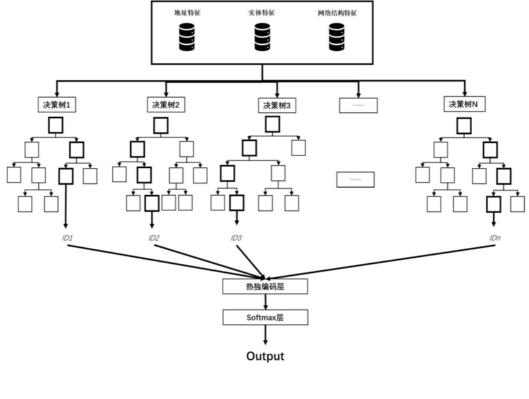


图3

联合特征向量构造单元 聚类单元 彩集单元

图4