(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 110309840 A
(43)申请公布日 2019.10.08

(21)申请号 201810258226.X

(22)申请日 2018.03.27

(71)申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 郑文豪 张雅淋 李龙飞

(74)专利代理机构 北京众达德权知识产权代理有限公司 11570
代理人 刘杰

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

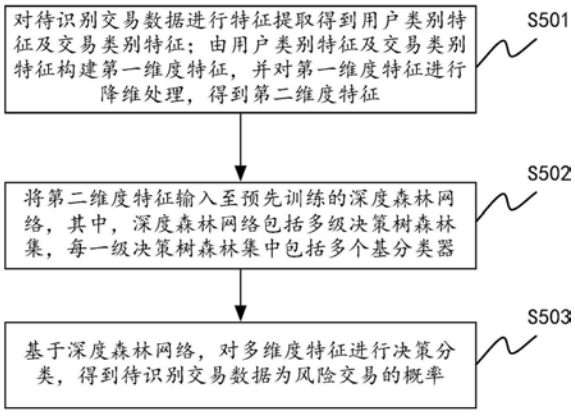
权利要求书3页 说明书9页 附图5页

(54)发明名称

风险交易识别方法、装置、服务器及存储介质

(57)摘要

本说明书实施例提供了一种风险交易识别方法，通过对交易数据的特征进行降维处理，针对降维后的特征，利用深度森林网络中每级决策树森林集的多个基分类器进行决策分类，最终确定出是否为风险交易的概率。本发明实施例通过对特征降维处理，可以防止过拟合或达到尽可能保留特征属性的效果。



CN 110309840 A

1.一种风险交易识别方法,包括:

对待识别交易数据进行特征提取,得到用户类别特征及交易类别特征;

由所述用户类别特征及交易类别特征组合得到第一维度特征,并对所述第一维度特征进行降维处理,得到第二维度特征;

将所述第二维度特征输入至预先训练的深度森林网络,其中,所述深度森林网络包括多级决策树森林集,每一级决策树森林集中包括多个基分类器;

基于所述深度森林网络,对多维度特征进行决策分类,得到待识别交易数据为风险交易的概率。

2.根据权利要求1所述的方法,还包括:基于交易样本训练出所述深度森林网络;

所述基于交易样本训练出所述深度森林网络包括:

收集有关风险交易的黑白样本,并对黑白样本数据进行特征提取得到第一维度特征,以及对第一维度特征进行降维处理,得到第二维度特征;

根据第二维度特征训练第一级决策树森林集的各个基分类器,并将前一级决策树森林集的输出特征与第二维度特征进行拼接,利用拼接特征训练下一级决策树森林集的各个基分类器;其中在每一级决策树森林集训练完成后判断是否达到预定结束条件,如果未达到才进行下一级决策树森林集的训练;

当达到预定结束条件时,结束训练,训练得到由多级决策树森林集构成的所述深度森林网络。

3.根据权利要求1或2所述的方法,所述对所述第一维度特征进行降维处理,得到第二维度特征包括:

根据特征类别,确定特征采样频率;

按照特征采样频率对所述初步维度特征进行采样,得到所述第二维度特征。

4.根据权利要求2所述的方法,所述基分类器包括一个或多个决策树;所述方法还包括:

根据黑白样本的比例,确定出决策树深度最大阈值;

设置所述基分类器中决策树的深度不超过所述深度最大阈值。

5.根据权利要求2所述的方法,

还包括:将黑白样本的数据划分为预置数目的分组;任选一个分组作为验证集,其余分组的数据集合作为训练集;

所述每级决策树森林集的训练过程中,是利用每个训练集分别训练每级决策树森林集中的各个基分类器的。

6.根据权利要求5所述的方法,还包括:

按照黑白样本比例,确定黑样本和白样本各自的样本采样频率;

按照黑白样各自的样本采样频率,分别对黑白样本进行采样,从而确保所述每个分组中黑白样本的数目相等或近似相等。

7.一种用于风险交易识别的深度森林网络的训练方法,包括:

收集有关风险交易的黑白样本,并对黑白样本数据进行特征提取得到得到用户类别特征及交易类别特征;由用户类别特征及交易类别特征构建得到第一维度特征,以及对第一维度特征进行降维处理,得到第二维度特征;

根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；

当达到预定结束条件时，结束训练，训练得到由多级决策树森林集构成的所述深度森林网络。

8.一种风险交易识别装置，包括：

特征提取及处理单元，用于对待识别交易数据进行特征提取得到得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对所述第一维度特征进行降维处理，得到第二维度特征；

预测单元，将所述第二维度特征输入至预先训练的深度森林网络，其中，所述深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器；基于所述深度森林网络，对多维度特征进行决策分类，得到待识别交易数据是否为风险交易的概率。

9.根据权利要求8所述的装置，还包括：网络训练单元；

所述网络训练单元包括：

样本获取子单元，用于收集有关风险交易的黑白样本；

特征提取及处理子单元，用于对黑白样本数据进行特征提取得到第一维度特征，以及对第一维度特征进行降维处理，得到第二维度特征；

训练执行子单元，用于根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；

网络确定子单元，用于当达到预定结束条件时，结束训练，得到由多级决策树森林集构成的所述深度森林网络。

10.根据权利要求8或9所述的装置，所述特征提取及处理单元或所述特征提取及处理子单元具体用于：根据特征类别，确定特征采样频率；按照特征采样频率对所述初步维度特征进行采样，得到所述第二维度特征。

11.根据权利要求9所述的装置，所述基分类器包括一个或多个决策树；所述网络训练单元还包括：

决策树深度控制子单元，用于根据黑白样本的比例，确定出决策树深度最大阈值；设置所述基分类器中决策树的深度不超过所述深度最大阈值。

12.根据权利要求9所述的装置，所述网络训练单元还包括：

样本分组子单元，用于将黑白样本的数据划分为预置数目的分组；任选一个分组作为验证集，其余分组的数据集合作为训练集；

所述训练执行子单元，在每级决策树森林集的训练过程中，是利用每个训练集分别训练每级决策树森林集中的各个基分类器的。

13.根据权利要求12所述的装置，所述网络训练单元还包括：

样本分组控制子单元，用于按照黑白样本比例，确定黑样本和白样本各自的样本采样频率；按照黑白样各自的样本采样频率，分别对黑白样本进行采样，从而确保所述每个分组

中黑白样本的数目相等或近似相等。

14.一种用于风险交易识别的深度森林网络的训练装置，包括：

样本获取单元，用于收集有关风险交易的黑白样本；

特征提取及处理单元，用于对黑白样本数据进行特征提取得到得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，以及对第一维度特征进行降维处理，得到第二维度特征；

训练执行单元，用于根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；

网络确定单元，用于当达到预定结束条件时，结束训练，得到由多级决策树森林集构成的所述深度森林网络。

15.一种服务器，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现权利要求1-11任一项所述方法的步骤。

16.一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现权利要求1-11任一项所述方法的步骤。

风险交易识别方法、装置、服务器及存储介质

技术领域

[0001] 本说明书实施例涉及互联网技术领域，尤其涉及一种风险交易识别方法、装置、服务器及存储介质。

背景技术

[0002] 随着互联网的快速发展，各种形式的业务不断涌现，如在线银行、在线支付、在线购物等基于互联网的服务业务。人们已经越来越习惯在网上进行各种生活或商务活动。

[0003] 由于互联网是一个开放的网络，任何人在任何地方都可以很方便地连接到互联网上。互联网在给人们生活提供便利的同时，也带来了风险。尤其是随着电子商务平台和第三方交易平台的发展，网络金融犯罪以及网上诈骗、信用卡盗刷等不断出现。因此，识别出风险交易越来越重要。

发明内容

[0004] 本说明书实施例提供及一种风险交易识别方法、装置、服务器及存储介质。

[0005] 第一方面，本说明书实施例提供一种风险交易识别方法，包括：对待识别交易数据进行特征提取得到得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对所述第一维度特征进行降维处理，得到第二维度特征；将所述第二维度特征输入至预先训练的深度森林网络，其中，所述深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器；基于所述深度森林网络，对多维度特征进行决策分类，得到待识别交易数据为风险交易的概率。

[0006] 第二方面，本说明书实施例提供一种用于风险交易识别的深度森林网络的训练方法，包括：收集有关风险交易的黑白样本，并对黑白样本数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，以及对第一维度特征进行降维处理，得到第二维度特征；根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；当达到预定结束条件时，结束训练，训练得到由多级决策树森林集构成的所述深度森林网络。

[0007] 第三方面，本说明书实施例提供一种风险交易识别装置，包括：特征提取及处理单元，用于对待识别交易数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对所述第一维度特征进行降维处理，得到第二维度特征；预测单元，将所述第二维度特征输入至预先训练的深度森林网络，其中，所述深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器；基于所述深度森林网络，对多维度特征进行决策分类，得到待识别交易数据为风险交易的概率。

[0008] 第四方面，本说明书实施例提供一种用于风险交易识别的深度森林网络的训练装置，包括：样本获取单元，用于收集有关风险交易的黑白样本；特征提取及处理单元，用于对

黑白样本数据进行特征提取得到用户类别特征及交易类别特征;由用户类别特征及交易类别特征构建第一维度特征,以及对第一维度特征进行降维处理,得到第二维度特征;训练执行单元,用于根据第二维度特征训练第一级决策树森林集的各个基分类器,并将前一级决策树森林集的输出特征与第二维度特征进行拼接,利用拼接特征训练下一级决策树森林集的各个基分类器;其中在每一级决策树森林集训练完成后判断是否达到预定结束条件,如果未达到才进行下一级决策树森林集的训练;网络确定单元,用于当达到预定结束条件时,结束训练,得到由多级决策树森林集构成的所述深度森林网络。

[0009] 第五方面,本说明书实施例提供一种服务器,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述任一项所述方法的步骤。

[0010] 第六方面,本说明书实施例提供一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现上述任一项所述方法的步骤。

[0011] 本说明书实施例有益效果如下:

[0012] 本发明实施例提出的风险交易识别方法中,通过对交易数据的特征进行降维处理,针对降维后的特征,利用深度森林网络中每级决策树森林集的多个基分类器进行决策分类,最终确定出是否为风险交易的概率。特别的,可根据特征类别确定采样频率,对于不同类型的特征,采用不同的采样方式,可以防止过拟合或达到尽可能保留特征属性的效果。另外,针对非法交易的样本有可能特别稀疏的情况,可采取黑白样本分别采样以及k折交叉验证的方式,保证在每一折中,正负比例都是一致或近似一致;还可以设置基分类器中决策树的深度不超过深度最大阈值,从而避免由于黑白样本比例过于悬殊导致的错分正常交易样本的问题。

附图说明

[0013] 图1为本说明书实施例风险交易识别的场景示意图;

[0014] 图2为本说明书实施例第一方面提供的用于风险交易识别的深度森林网络训练方法中深度森林网络示意图;

[0015] 图3为本说明书实施例第一方面提供的用于风险交易识别的深度森林网络训练方法中每个森林内部示意图;

[0016] 图4为本说明书实施例第一方面提供的用于风险交易识别的深度森林网络的训练方法流程图;

[0017] 图5为本说明书实施例第二方面提供的风险交易识别方法流程图;

[0018] 图6为本说明书实施例第三方面提供的用于风险交易识别的深度森林网络的训练装置结构示意图;

[0019] 图7为本说明书实施例第四方面提供的风险交易识别装置结构示意图;

[0020] 图8为本说明书实施例第五方面提供的服务器结构示意图。

具体实施方式

[0021] 为了更好的理解上述技术方案,下面通过附图以及具体实施例对本说明书实施例的技术方案做详细的说明,应当理解本说明书实施例以及实施例中的具体特征是对本说明

书实施例技术方案的详细的说明，而不是对本说明书技术方案的限定，在不冲突的情况下，本说明书实施例以及实施例中的技术特征可以相互组合。

[0022] 请参见图1，为本说明书实施例的风险交易(异常交易、非法交易)识别的场景示意图。终端100位于用户侧，与网络侧的服务器200通信。终端100中的交易处理客户端101可以是基于互联网实现业务的APP或网站，为用户提供交易的界面并将交易数据提供给网络侧进行处理；服务器200利用预先训练的深度森林网络201用于对交易处理客户端101中涉及的风险交易进行识别。

[0023] 随着人工智能的兴起，机器学习作为人工智能中最重要的技术，也日益受到人们的重视。机器学习算法具有更加灵活更为智能的优点。如今的基于多种类型特征(离散特征，连续特征，类别属性特征)的机器学习解决方案，大都采用梯度决策提升树，然而这种结构虽然能够适应多种场景，但也有一定的局限性，例如对于黑样本较少的场景，如何找到更多的黑用户，如何提高预测准确性等方面还存在不足。

[0024] 在风控场景中，通常需要找到带有安全隐患的交易，这类交易称为非法交易，相较于正常交易而言，这类交易的数目非常少，通常是一比几百甚至几千，而且异常交易有着形形色色的区别，因而挖掘非法交易是一件较为困难的事情。对此，本发明实施例提出一种用于风险交易识别的深度森林网络训练方法以及风险交易识别方法，应用在智能风控领域，通过特征降维、样本采样、限定决策树深度等方式，能够比以往的算法找到更多的非法交易。

[0025] 深度森林网络，是借鉴集成学习的思想，以基于决策树的集合(森林)为基分类器，构建一个多层(多级)网络，网络的层数可以自适应得到。每层网络的节点为一个梯度提升决策树。

[0026] 如图2所示，为深度森林网络的一个示意图。该深度森林网络包括L级(L层)，每个级是决策树森林的一个集合(决策树森林集)，即集成的集成(ensemble of ensembles)。每一级决策树森林集包括多个基分类器(森林:forest)。每一级决策树森林集可以包含不同类型的森林(例如随机森林或完全随机树木森林)，从而提高网络的多样性。例如，图2中级决策树森林集包括四个基分类器。每个基分类器又由一个或多个决策树构成的。如图3所示，示出了一个基分类器内部包括三个决策树的情况。

[0027] 在网络训练或预测过程中，级联的各级决策树森林集中，除了最后一级决策树森林集中的每一级决策树森林集的输入，都是前一级处理的特征信息与原始特征拼接的拼接特征(其中，第一级决策树森林集没有前一级，因此输入仅是原始特征)。

[0028] 图2中，level-1的输入是原始特征，假设是二分类(有两个类要预测)问题，则level-1每个基分类器将输出二维类向量(class vector)，则四个基分类器每一个都将产生一个二维的类向量，得到八维的类向量(4×2)；继而，在level-2，将该八维类向量与原始特征向量相拼接，将接收(n×c+d)个增强特征(augmented feature)，其中，d为输入的初始特征的数量、n为基分类器个数、c为分类数目；同理在level-3至level-(L-1)均与level-2类似处理；在最后一级level-L，输入仅是上一级的输出(并不拼接原始特征)，输出得到八维类向量；最后对这八维类向量进行平均取值等处理，最终输出二分类的二维类向量。

[0029] 第一方面，本说明书实施例提供一种用于风险交易识别的深度森林网络训练方法流程图方法，请参考图4，包括步骤S401-S406。

[0030] S401:收集有关风险交易的黑白样本,并对黑白样本数据进行特征提取得到得到用户类别特征及交易类别特征,并根据用户类别特征及交易类别特征构建第一维度特征。

[0031] 从历史交易中,分别收集交易的黑白样本。黑样本是指风险交易样本,白样本是指正常交易的样本。在实际场景中,风险交易相较于正常交易毕竟是少数,因此往往存在黑样本数量不足而导致训练到的网络准确度不高的问题。对此本发明实施例采取多种方式(特征采样、样本采样、限定决策树深度等)进行改进,后续有说明。

[0032] 得到黑白样本之后,对黑白样本进行特征提取,得到包括多个类别的特征。例如得到:用户类别特征(例如:性别,年龄,历史交易笔数)以及交易类别特征(例如,成交量,交易额度,频次)等。将所有特征采样向量方式进行表示,得到第一维度特征。一般情况下,第一维度特征维度数目庞大,例如是一个几百维度的特征向量。如果直接将该第一维度特征输入到深度神经网络进行训练,则势必会降低网络训练效率;而且考虑到拟合效果,本发明实施例在S402中,第一维度特征进行降维处理。

[0033] S402:对第一维度特征进行降维处理,得到第二维度特征。

[0034] 为了方便,下面将“第二维度特征”表示为“d维特征”。

[0035] 如前述参考图2介绍的,在第i级(i小于最大级数L)每一级的输入都是(d+n*c)维特征,在非法交易场景中,通常都是二分类问题(c=2),当d很大时,n*c的值相对于d而言会微不足道,这样特征提供的信息较少,对后续层数的拟合则较差。因此为了提高拟合效果,当d远大于n*c时,可对初始得到的第一维度特征进行降维处理。

[0036] 在一种可选方式中,对第一维度特征进行降维处理的具体方式为:根据特征类别,确定特征采样频率,确定特征采样频率;按照特征采样频率对初步维度特征进行采样,得到第二维度特征。

[0037] 之所以根据特征类别确定采样频率,是因为在非法交易中,一个样本包含着不同类别的特征,不同的特征之下样本密度也是不同的,因此对于不同类型的特征,采用不同的采样方式。例如:对于用户类别特征(例如:性别,年龄,历史交易笔数),通常都是稀疏的,因此可以采用较低的采样频率,这样不仅能防止过拟合,还能保证在训练时,缺失值过多导致拟合难度增加。对于交易类别特征(例如,成交量,交易额度,频次),这些特征呈现伽马分布,都是连续值特征,缺失值较少,因此,可采取较高的采样频率,尽可能保留这些特征属性。

[0038] S403:根据第二维度特征训练第一级决策树森林集的各个基分类器。

[0039] S404:将前一级决策树森林集的输出特征与第二维度特征进行拼接,利用拼接特征训练下一级决策树森林集的各个基分类器;

[0040] S405:判断是否达到预定结束条件,如果未达到结束条件,则返回重复执行步骤S404;如果达到结束条件,则执行步骤S406。

[0041] 其中,结束条件可以包括多个,例如,当预测结果在评价指标上没有提升,或是已经达到了最大网络层数(级数L)则确定达到了结束条件。

[0042] S406:当达到预定结束条件时,结束训练,训练得到由多级决策树森林集构成的深度森林网络。

[0043] 上述步骤S403-S406描述了对每级决策树森林集的基分类器进行训练从而得到深度森林网络的过程;具体可参考上述图2及相关描述。例如参考图2:level-1的输入是原始

特征，假设是二分类(有两个类要预测)问题，则level-1每个基分类器将输出二维类向量(class vector)，则四个基分类器每一个都将产生一个二维的类向量，得到八维的类向量(4×2)；继而，在level-2，将该八维类向量与原始特征向量相拼接，将接收(n×c+d)个增强特征(augmented feature)，其中，d为输入的初始特征的数量、n为基分类器个数、c为分类数目；同理在level-3至level-(L-1)均与level-2类似处理；在最后一级level-L，输入仅是上一级的输出(并不拼接原始特征)，输出得到八维类向量；最后对这八维类向量进行平均取值等处理，最终输出二分类的二维类向量。

[0044] 如前提到的，在实际场景中，风险交易相较于正常交易毕竟是少数，因此往往存在黑样本数量不足而导致训练到的网络准确度不高的问题。为此，考虑到黑样本数量相比于白样本数量很少的情况，且通过简单划分即可区分非法交易和正常交易的条件下(由于非法交易的行为模式应当与正常行为相距甚远)，可以通过限定决策树最大深度的方式，一定程度减轻黑样本过少而导致的训练不准确的问题。

[0045] 因此，在一种可选方式中，可根据黑白样本的比例，确定出决策树深度最大阈值；设置基分类器中决策树的深度不超过深度最大阈值。例如，每个基分类器的树深度不超过5，因为，只需要简单的划分就可以区分非法交易和正常交易，如果树的深度太深，则容易错分正常交易样本，因为正常样本之间的模式并不是完全一致的。

[0046] 假定在一个例子中，对于d维特征，是c分类问题；假定有n个基分类器(可参考图2中每级决策树森林集包括4个基分类器)；每个基分类器包括一个或多个决策树。

[0047] 对于黑白样本比例不均(黑样本太少)的问题，可在对每级决策树森林集的各个基分类器进行训练之前，对输入的样本进行划分，通过k折交叉验证的方式进行，以缓解过拟合问题。

[0048] 交叉验证，是将原始数据(dataset)进行分组，一部分做为训练集(train set)，另一部分做为验证集(validation set or test set)，首先用训练集对分类器进行训练，再利用验证集来测试训练得到的模型(model)，以此来做为评价分类器的性能指标。

[0049] 在一种可选方式中，对黑白样本进行如下预处理：将黑白样本的数据划分为预置数目的分组；任选一个分组作为验证集，其余分组的数据集合作为训练集；在每级决策树森林集的训练过程中，是利用每个训练集分别训练每级决策树森林集中的各个基分类器的。其中，按照黑白样本比例，确定黑样本和白样本各自的样本采样频率；按照黑白样各自的样本采样频率，分别对黑白样本进行采样，从而确保每个分组中黑白样本的数目相等或近似相等。

[0050] 例如，假设黑样本有100个，白样本有500个，按照黑白样本比例，设定黑样本采样频率为1/2、白样本采样频率为1/10，则采样出50个黑样本及50个白样本。将选取出的总计100个黑白样本进行随机划分分为三个分组：分组1、分组2、分组3；则得到三种组合方式：分组1为验证集，分组2、3的数据集合作为训练集；分组2为验证集，分组1、3的数据集合作为训练集；分组3为验证集，分组1、2的数据集合作为训练集。在训练基分类器过程中，需要针对上述三个训练集分别进行训练。

[0051] 这样处理的目的在于，非法交易的样本有可能特别稀疏，有可能某一折中，黑样本太少，而引入偏差，因而可采取黑白样本分别采样的方式，保证在每一折中，正负比例都是一致或近似一致。

[0052] 第二方面，本说明书实施例提供一种风险交易识别方法。参考图5，该方法包括：

[0053] S501：对待识别交易数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对第一维度特征进行降维处理，得到第二维度特征；

[0054] S502：将第二维度特征输入至预先训练的深度森林网络，其中，深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器；

[0055] S503：基于深度森林网络，对多维度特征进行决策分类，得到待识别交易数据为风险交易的概率。

[0056] 其中对于特征降维处理以及深度森林网络的训练过程请参见本发明实施例前述内容。

[0057] 假定深度森林网络为L层(即包括L级决策树森林集)，则在利用深度森林网络对待识别交易数据进行预测的过程中，执行如下过程：

[0058] (1)利用n个分类器对待识别交易数据进行预测：对于待识别交易数据，得到n*c个预测结果，拼接到原始的d维特征上，得到新的(d+n*c)维特征；

[0059] (2)对于最后L层的n*c个预测结果，在n个分类结果取平均，得到最终的c个预测结果，在c个预测结果便是深度森林网络在c个类别上的最终预测概率。

[0060] 可见，本发明实施例提出的风险交易识别方法中，通过对交易数据的特征进行降维处理，针对降维后的特征，利用深度森林网络中每级决策树森林集的多个基分类器进行决策分类，最终确定出风险交易的概率。特别的，可根据特征类别确定采样频率，对于不同类型的特征，采用不同的采样方式，可以防止过拟合或达到尽可能保留特征属性的效果。另外，针对非法交易的样本有可能特别稀疏的情况，可采取黑白样本分别采样以及k折交叉验证的方式，保证在每一折中，正负比例都是一致或近似一致；还可以设置基分类器中决策树的深度不超过深度最大阈值，从而避免由于黑白样本比例过于悬殊导致的错分正常交易样本的问题。

[0061] 第三方面，基于同一发明构思，本说明书实施例提供一种用于风险交易识别的深度森林网络的训练装置，请参考图6，包括：

[0062] 样本获取单元601，用于收集有关风险交易的黑白样本；

[0063] 特征提取及处理单元602，用于对黑白样本数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，以及对第一维度特征进行降维处理，得到第二维度特征；

[0064] 训练执行单元603，用于根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；

[0065] 网络确定单元604，用于当达到预定结束条件时，结束训练，得到由多级决策树森林集构成的所述深度森林网络。

[0066] 在一种可选方式中，所述特征提取及处理单元602具体用于：根据特征类别，确定特征采样频率；按照特征采样频率对所述初步维度特征进行采样，得到所述第二维度特征。

[0067] 在一种可选方式中，所述基分类器包括一个或多个决策树；所述装置还包括：

[0068] 决策树深度控制单元605，用于根据黑白样本的比例，确定出决策树深度最大阈值；设置所述基分类器中决策树的深度不超过所述深度最大阈值。

[0069] 在一种可选方式中，所述装置还包括：

[0070] 样本分组单元606，用于将黑白样本的数据划分为预置数目的分组；任选一个分组作为验证集，其余分组的数据集合作为训练集；

[0071] 所述训练执行单元603，在每级决策树森林集的训练过程中，是利用每个训练集分别训练每级决策树森林集中的各个基分类器的。

[0072] 在一种可选方式中，所述装置还包括：

[0073] 样本分组控制单元607，用于按照黑白样本比例，确定黑样本和白样本各自的样本采样频率；按照黑白样各自的样本采样频率，分别对黑白样本进行采样，从而确保所述每个分组中黑白样本的数目相等或近似相等。

[0074] 第四方面，基于同一发明构思，本说明书实施例提供一种风险交易识别装置，请参考图7，包括：

[0075] 特征提取及处理单元701，用于对待识别交易数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对所述第一维度特征进行降维处理，得到第二维度特征；

[0076] 预测单元702，将所述第二维度特征输入至预先训练的深度森林网络，其中，所述深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器；基于深度森林网络，对所述多维度特征进行决策分类，得到待识别交易数据为风险交易的概率。

[0077] 在一种可选方式中，还包括：网络训练单元703；

[0078] 所述网络训练单元703包括：

[0079] 样本获取子单元7031，用于收集有关风险交易的黑白样本；

[0080] 特征提取及处理子单元7032，用于对黑白样本数据进行特征提取得到第一维度特征，以及对第一维度特征进行降维处理，得到第二维度特征；

[0081] 训练执行子单元7033，用于根据第二维度特征训练第一级决策树森林集的各个基分类器，并将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器；其中在每一级决策树森林集训练完成后判断是否达到预定结束条件，如果未达到才进行下一级决策树森林集的训练；

[0082] 网络确定子单元7034，用于当达到预定结束条件时，结束训练，得到由多级决策树森林集构成的所述深度森林网络。

[0083] 在一种可选方式中，所述特征提取及处理单元702或所述特征提取及处理子单元7032具体用于：根据特征类别，确定特征采样频率；按照特征采样频率对所述初步维度特征进行采样，得到所述第二维度特征。

[0084] 在一种可选方式中，所述基分类器包括一个或多个决策树；所述网络训练单元703还包括：

[0085] 决策树深度控制子单元7035，用于根据黑白样本的比例，确定出决策树深度最大阈值；设置所述基分类器中决策树的深度不超过所述深度最大阈值。

[0086] 在一种可选方式中，所述网络训练单元703还包括：

[0087] 样本分组子单元7036，用于将黑白样本的数据划分为预置数目的分组；任选一个

分组作为验证集，其余分组的数据集合作为训练集；

[0088] 所述训练执行子单元7034，在每级决策树森林集的训练过程中，是利用每个训练集分别训练每级决策树森林集中的各个基分类器的。

[0089] 在一种可选方式中，所述网络训练单元703还包括：

[0090] 样本分组控制子单元7037，用于按照黑白样本比例，确定黑样本和白样本各自的样本采样频率；按照黑白样各自的样本采样频率，分别对黑白样本进行采样，从而确保所述每个分组中黑白样本的数目相等或近似相等。

[0091] 第五方面，基于与前述实施例中风险交易识别方法或用于风险交易识别的深度森林网络的学习方法同样的发明构思，本发明还提供一种服务器，如图8所示，包括存储器804、处理器802及存储在存储器804上并可在处理器802上运行的计算机程序，所述处理器802执行所述程序时实现前文所述风险交易识别方法或用于风险交易识别的深度森林网络的学习方法的步骤。

[0092] 其中，在图8中，总线架构(用总线800来代表)，总线800可以包括任意数量的互联的总线和桥，总线800将包括由处理器802代表的一个或多个处理器和存储器804代表的存储器的各种电路链接在一起。总线800还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起，这些都是本领域所公知的，因此，本文不再对其进行进一步描述。总线接口806在总线800和接收器801和发送器803之间提供接口。接收器801和发送器803可以是同一个元件，即收发机，提供用于在传输介质上与各种其他装置通信的单元。处理器802负责管理总线800和通常的处理，而存储器804可以被用于存储处理器802在执行操作时所使用的数据。

[0093] 第六方面，基于与前述实施例中风险交易识别方法或用于风险交易识别的深度森林网络的学习方法的发明构思，本发明还提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前文所述风险交易识别方法或用于风险交易识别的深度森林网络的学习方法的步骤。

[0094] 本说明书是参照根据本说明书实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的设备。

[0095] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令设备的制造品，该指令设备实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0096] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0097] 尽管已描述了本说明书的优选实施例，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说明书范围的所有变更和修改。

[0098] 显然，本领域的技术人员可以对本说明书进行各种改动和变型而不脱离本说明书的精神和范围。这样，倘若本说明书的这些修改和变型属于本说明书权利要求及其等同技术的范围之内，则本说明书也意图包含这些改动和变型在内。

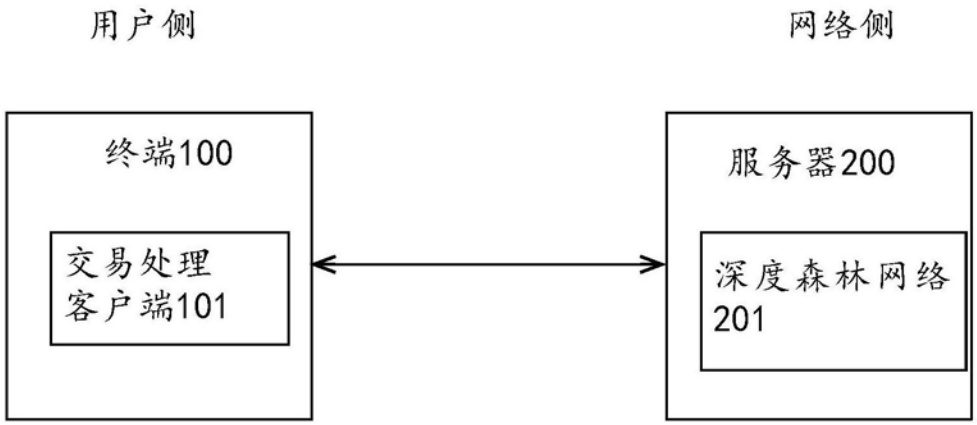

用户侧
网络侧
终端100
交易处理
客户端101
服务器200
深度森林网络
201

图1

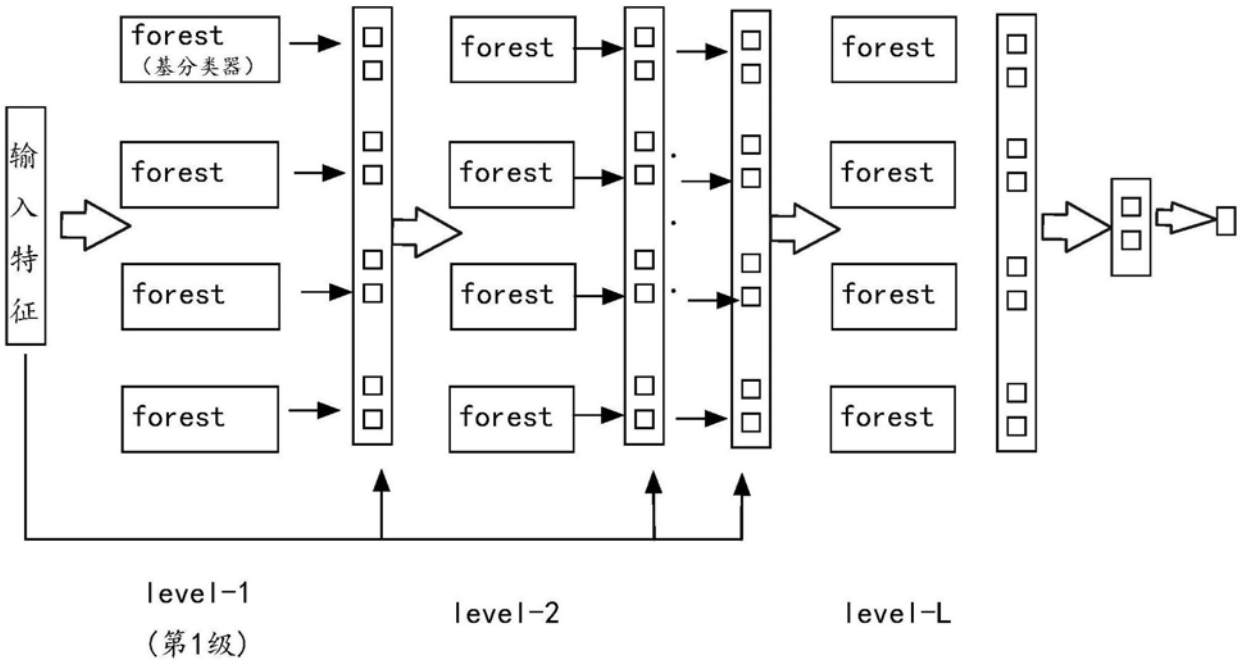

输入特征
forest
（基分类器）
forest
forest
forest
forest
forest
forest
forest
forest
forest
forest
forest
forest
level-1
（第1级）
level-2
level-L

图2

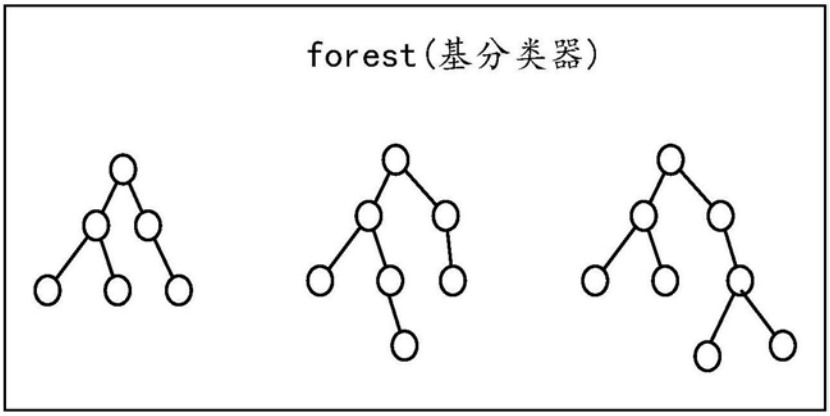


图3

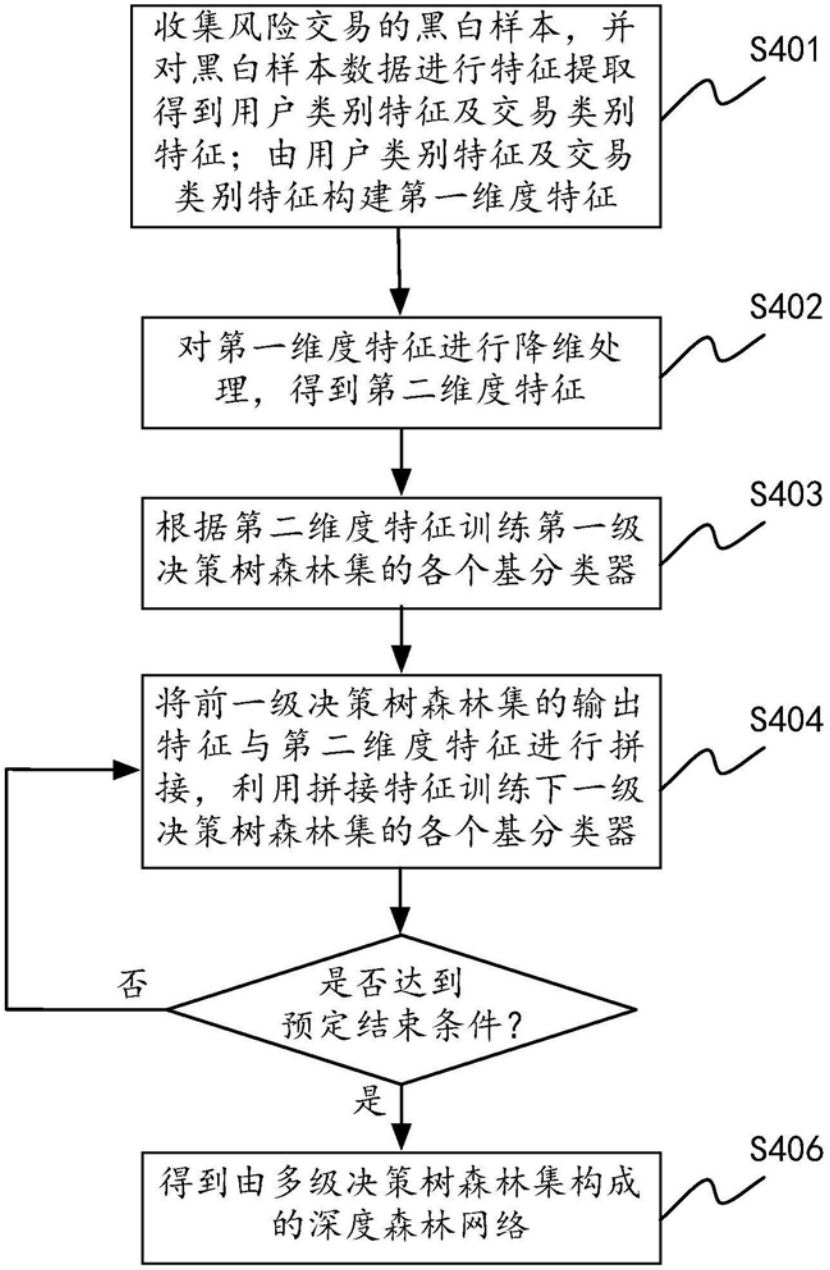

收集风险交易的黑白样本，并对黑白样本数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征
S401
对第一维度特征进行降维处理，得到第二维度特征
S402
根据第二维度特征训练第一级决策树森林集的各个基分类器
S403
将前一级决策树森林集的输出特征与第二维度特征进行拼接，利用拼接特征训练下一级决策树森林集的各个基分类器
S404
否
是否达到预定结束条件？
是
得到由多级决策树森林集构成的深度森林网络
S406

图4

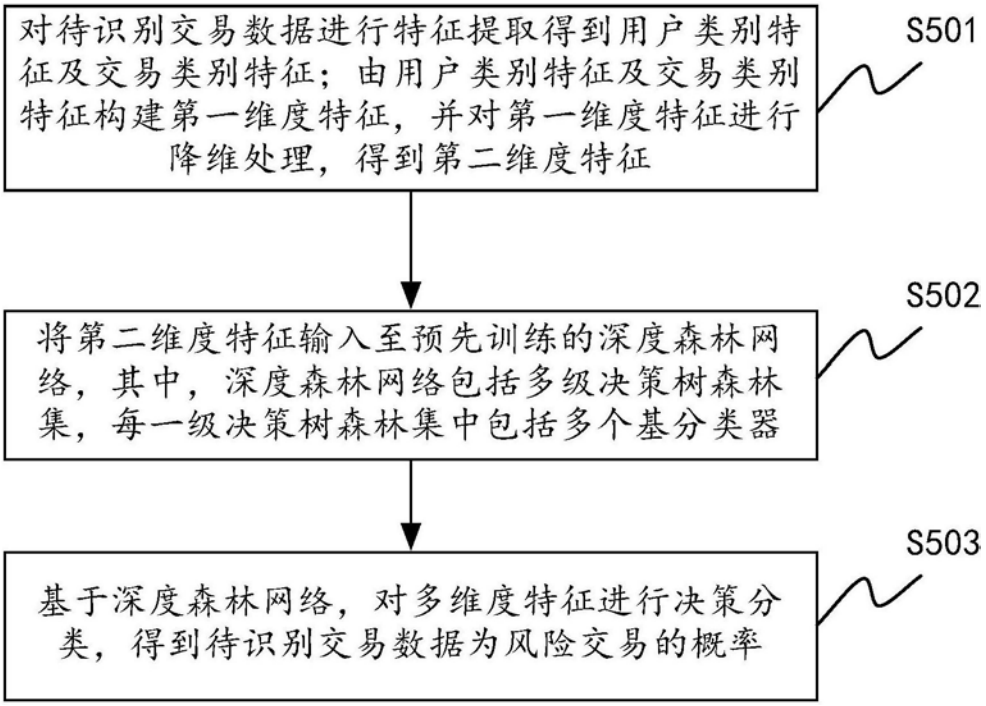

对待识别交易数据进行特征提取得到用户类别特征及交易类别特征；由用户类别特征及交易类别特征构建第一维度特征，并对第一维度特征进行降维处理，得到第二维度特征
S501
将第二维度特征输入至预先训练的深度森林网络，其中，深度森林网络包括多级决策树森林集，每一级决策树森林集中包括多个基分类器
S502
基于深度森林网络，对多维度特征进行决策分类，得到待识别交易数据为风险交易的概率
S503

图5

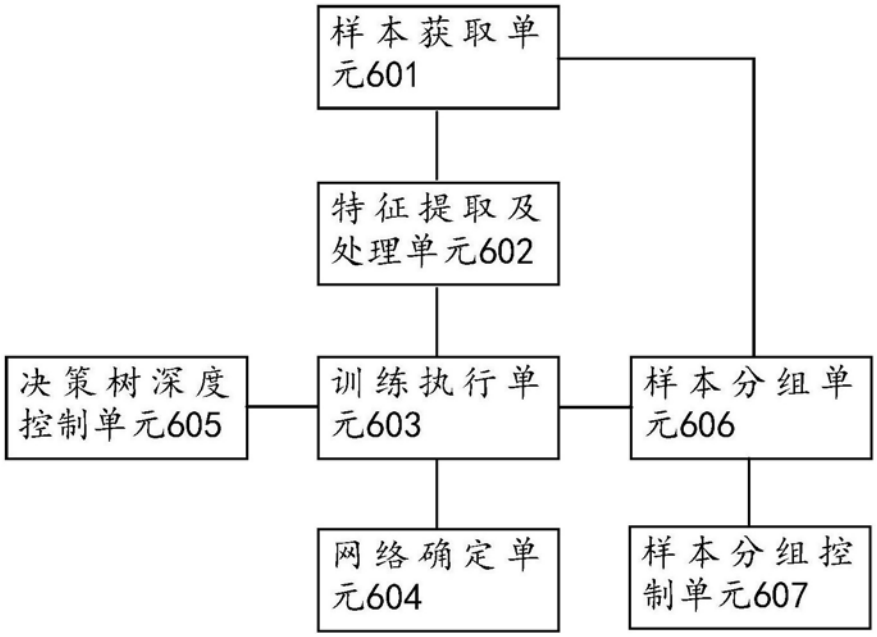

样本获取单元601
特征提取及处理单元602
决策树深度控制单元605
训练执行单元603
样本分组单元606
网络确定单元604
样本分组控制单元607

图6

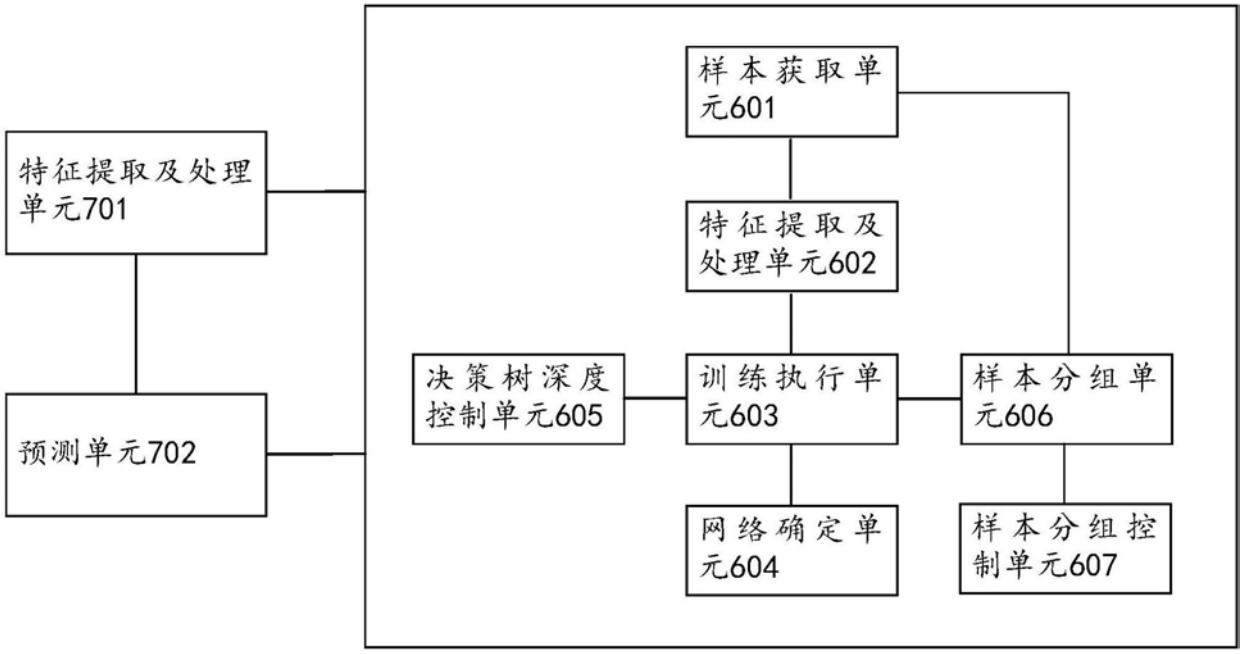

样本获取单元601
特征提取及处理单元701
特征提取及处理单元602
决策树深度控制单元605
训练执行单元603
样本分组单元606
预测单元702
网络确定单元604
样本分组控制单元607

图7

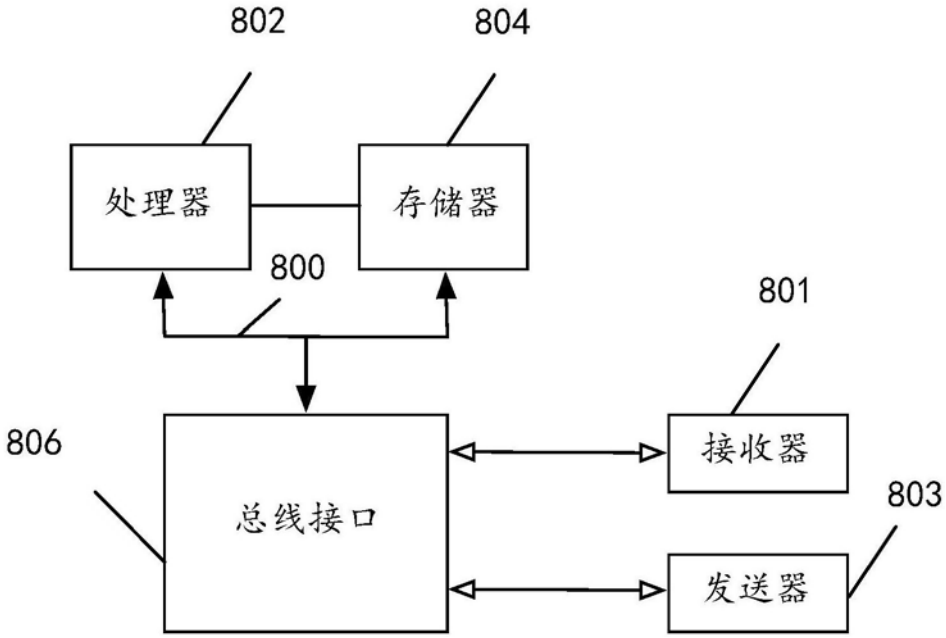

802
804
处理器
存储器
800
801
806
总线接口
接收器
803
发送器

图8