



## (12)发明专利申请

(10)申请公布号 CN 107944874 A

(43)申请公布日 2018.04.20

(21)申请号 201711331628.X

(22)申请日 2017.12.13

(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 陈明星

(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.

G06Q 20/40(2012.01)

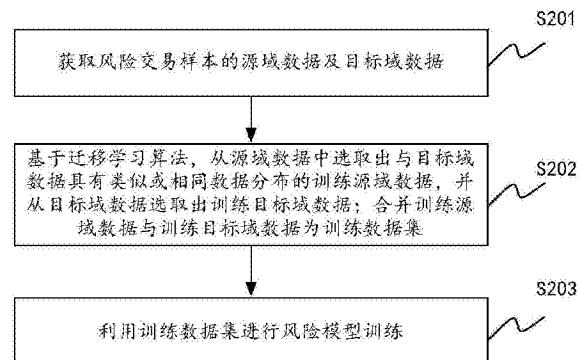
权利要求书3页 说明书7页 附图3页

(54)发明名称

基于迁移学习的风控方法、装置及系统

(57)摘要

本说明书实施例提供一种基于迁移学习的风控方法，基于迁移学习算法，从风险交易样本的源域数据中选取训练源域数据作为目标域数据的有效补充，在目标域数据不足等情况下，有助于打通不同数据源，提升风控效率。



1. 一种基于迁移学习的风控方法,包括:

获取风险交易样本的源域数据及目标域数据;

基于迁移学习算法,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据,并从所述目标域数据选取训练目标域数据;合并所述训练源域数据与所述训练目标域数据为训练数据集;

利用所述训练数据集进行风险模型训练。

2. 根据权利要求1所述的方法,在获取源域数据及目标域数据之后,还包括:

合并所述源域数据及目标域数据为具有相同数据结构的初始合并数据集;其中,所述训练数据集是从所述初始合并数据集中选定的。

3. 根据权利要求1所述的方法,所述从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据包括:

基于数据特征提取及特征聚类,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

基于深度学习对数据建模的模型参数分布相似度,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

基于数据迭代分类过程中权重动态更新,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据。

4. 根据权利要求1所述的方法,所述从所述目标域数据选取训练目标域数据包括:

根据所述目标域数据中各数据的完整性,选取训练目标域数据。

5. 根据权利要求3所述的方法,所述基于数据迭代分类过程中权重动态更新,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据,包括:

为所述源域数据中每一个样本数据赋予权重;

在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则降低所述样本数据的权重;

根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据。

6. 根据权利要求5所述的方法,所述从所述目标域数据选取训练目标域数据包括:

为所述目标域数据中每一个样本数据赋予权重;

在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则提高所述样本数据的权重;

根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从所述目标域数据选取训练目标域数据。

7. 根据权利要求1-6任一项所述的方法,所述利用所述训练数据集进行风险分类模型训练包括:

针对所述训练数据集,通过二分类算法进行训练,得到风险识别模型。

8. 根据权利要求7所述的方法,在得到风险识别模型之后,还包括:

根据所述风险识别模型,对业务数据进行识别,确定出业务风险分值;

对于风险分值高于风险阈值的业务进行控制。

9. 一种基于迁移学习的风控装置,包括:

数据获取单元,用于获取风险交易样本的源域数据及目标域数据;

训练数据集确定单元,用于基于迁移学习算法,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据,并从所述目标域数据选取训练目标域数据;合并所述训练源域数据与所述训练目标域数据为训练数据集;

训练单元,用于利用所述训练数据集进行风险模型训练。

10. 根据权利要求9所述的装置,还包括:

初始合并单元,用于合并所述源域数据及目标域数据为具有相同数据结构的初始合并数据集;其中,所述训练数据集是从所述初始合并数据集中选定的。

11. 根据权利要求9所述的装置,所述训练数据集确定单元具体用于:

基于数据特征提取及特征聚类,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

基于数据深度学习建模及模型参数分布相似度,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

基于数据迭代分类过程中权重动态更新,从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据。

12. 根据权利要求9所述的装置,所述训练数据集确定单元具体用于:根据所述目标域数据中各数据的完整性,选取训练目标域数据。

13. 根据权利要求11所述的装置,所述训练数据集确定单元包括:

源域权重赋予子单元,用于为所述源域数据中每一个样本数据赋予权重;

源域权重动态更新子单元,用于在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则降低所述样本数据的权重;

源域数据确定子单元,用于根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据。

14. 根据权利要求13所述的装置,所述训练数据集确定单元还包括:

目标域权重赋予子单元,用于为所述目标域数据中每一个样本数据赋予权重;

目标域权重动态更新子单元,用于在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则提高所述样本数据的权重;

目标域数据确定子单元,用于根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从所述目标域数据选取训练目标域数据。

15. 根据权利要求9-14任一项所述的装置,所述训练单元具体用于:针对所述训练数据集,通过二分类算法进行训练,得到风险识别模型。

16. 根据权利要求15所述的装置,还包括:

风险识别单元,用于根据所述风险识别模型,对业务数据进行识别,确定出业务风险分值;

风险控制单元,用于对于风险分值高于风险阈值的业务进行控制。

17. 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1-8任一项所述方法的步骤。

18. 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1-8任一项所述方法的步骤。

## 基于迁移学习的风控方法、装置及系统

### 技术领域

[0001] 本说明书实施例涉及互联网技术领域，尤其涉及一种基于迁移学习的风控方法、装置及系统。

### 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。人们已经接受并越来越习惯在网上进行各种商务及生活活动。

[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欺诈、信用卡盗刷等不断出现。因此，对交易进行风险识别及控制越来越重要。

### 发明内容

[0004] 本说明书实施例提供及一种基于迁移学习的风控方法、装置及系统。

[0005] 第一方面，本说明书实施例提供一种基于迁移学习的风控方法，包括：

[0006] 获取风险交易样本的源域数据及目标域数据；

[0007] 基于迁移学习算法，从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据，并从所述目标域数据选取训练目标域数据；合并所述训练源域数据与所述训练目标域数据为训练数据集；

[0008] 利用所述训练数据集进行风险模型训练。

[0009] 第二方面，本说明书实施例提供一种基于迁移学习的风控装置，包括：

[0010] 数据获取单元，用于获取风险交易样本的源域数据及目标域数据；

[0011] 训练数据集确定单元，用于基于迁移学习算法，从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据，并从所述目标域数据选取训练目标域数据；合并所述训练源域数据与所述训练目标域数据为训练数据集；

[0012] 训练单元，用于利用所述训练数据集进行风险模型训练。

[0013] 第三方面，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，所述处理器执行上述任一项所述方法的步骤。

[0014] 第四方面，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。

[0015] 本说明书实施例有益效果如下：

[0016] 在本说明书实施例提供的基于迁移学习的风控方法中，基于迁移学习算法，从风险交易样本的源域数据中选取训练源域数据作为目标域数据的有效补充，在目标域数据不足等情况下，有助于模型建立的准确性。通过打通不同数据源（例如不同国家数据或不同市场数据），提升风控效率。

## 附图说明

- [0017] 图1为本说明书实施基于迁移学习进行风控例的场景示意图；
- [0018] 图2为本说明书实施例第一方面提供的基于迁移学习的风控方法流程图；
- [0019] 图3为本说明书实施例第一方面提供的基于迁移学习的风控方法实例整体示意图；
- [0020] 图4为本说明书实施例第一方面提供的基于迁移学习的风控方法实例迁移学习算法逻辑示意图；
- [0021] 图5为本说明书实施例第二方面提供的基于迁移学习的风控装置结构示意图；
- [0022] 图6为本说明书实施例第三方面提供的基于迁移学习的风控服务器结构示意图。

## 具体实施方式

[0023] 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。

[0024] 请参见图1，为本说明书实施例的基于迁移学习进行风控例的场景示意图。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中的异常交易识别系统201用于对交易处理客户端101中涉及的异常交易进行识别以及风险控制。

[0025] 为了对交易进行风险识别，需要利用风险交易样本数据进行风险识别模型建立。传统的机器学习的模型都是建立在训练数据和测试数据服从相同的数据分布的假设上。典型的比如有监督学习，可以在训练数据上面训练得到一个分类器，用于测试数据。但是在许多的情况下，这种同分布的假设并不满足，比如历史数据和最近的数据在风控场景中随着案件形势的不同会导致不同的数据分布的，用历史数据训练的模型在最近的风控识别上性能会下降。如果用新数据重新进行建模，首先就缺乏标注样本，因为新数据的标注又是费时费力。加上丢弃历史的标注数据又是十分可惜的，所以如何利用好这些历史的标注数据，来帮忙在新数据上进行建模显得至关重要。

[0026] 第一方面，本说明书实施例提供一种基于迁移学习的风控方法，请参考图2，包括S201-S203。

[0027] S201：获取风险交易样本的源域数据及目标域数据。

[0028] 为了对交易进行风险识别，需要利用风险交易样本数据进行风险识别模型建立。源域数据 (source domain) 指代原有成熟市场的历史数据样本，目标域数据 (target domain) 指代新市场的新数据样本。一般来说，target domain的数据量级和案件数都远小于source domain。

[0029] 在一种可选方式中在获取源域数据及目标域数据之后，还可执行如下步骤：合并源域数据及目标域数据为具有相同数据结构的初始合并数据集；其中，后续的训练数据集是从初始合并数据集中选定的。

[0030] S202:基于迁移学习算法,从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据,并从目标域数据选取出训练目标域数据;合并训练源域数据与训练目标域数据为训练数据集。

[0031] 在传统的机器学习的框架下,学习的任务就是在给定充分训练数据的基础上来学习一个分类模型;然后利用这个学习到的模型来对数据进行分类与预测。

[0032] 然而,机器学习算法在当前的Web挖掘研究中存在着一个关键的问题:一些新出现的领域中的大量训练数据非常难得到。传统的机器学习需要对每个领域都标定大量训练数据,这将会耗费大量的人力与物力。而没有大量的标注数据,会使得很多与学习相关研究与应用无法开展。其次,传统的机器学习假设训练数据与测试数据服从相同的数据分布。然而,在许多情况下,这种同分布假设并不满足。通常可能发生的情况如训练数据过期。这往往需要去重新标注大量的训练数据以满足训练的需要,但标注新数据是非常昂贵的,需要大量的人力与物力。从另外一个角度上看,如果有了大量的、在不同分布下的训练数据,完全丢弃这些数据也是非常浪费的。

[0033] 如何合理的利用这些历史数据就是迁移学习主要解决的问题。迁移学习可以从现有的历史数据中迁移知识,用来帮助将来的学习。迁移学习(Transfer Learning)的目标是将从一个环境中学到的知识用来帮助新环境中的学习任务。因此,迁移学习不会像传统机器学习那样作同分布假设。

[0034] 在本说明书实施例中,“从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据”的方式包括但不限于:

[0035] (1) 基于数据特征提取及特征聚类,从所述源域数据中选取出与所述目标域数据具有类似或相同数据分布的训练源域数据。

[0036] 有多种学习算法可实现基于特征的迁移学习,包括基于特征的有监督迁移学习与基于特征的无监督迁移学习,例如CoCC算法、TPLSA算法、谱分析算法、自学习算法等。其中利用互聚类算法产生一个公共的特征表示,从而帮助迁移学习算法。其思想是使用互聚类算法同时对源域数据与目标域数据数据进行聚类,得到一个共同的特征表示,这个新的特征表示优于只基于源域数据的特征表示。通过把源域数据表示在这个新的空间里,以实现迁移学习。

[0037] (2) 基于深度学习对数据建模的模型参数分布相似度,从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据。

[0038] 在利用源域数据及目标域数据进行基于深度学习建模过程中,不断优化模型参数;将根据模型参数分布相似度,确定出与目标域数据具有类似或相同数据分布的训练源域数据。

[0039] (3) 基于数据迭代分类过程中权重动态更新,从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据。

[0040] 基于实例的迁移学习,尽管目标域数据和源域数据或多或少会有些不同,但是源域数据中应该还是会存在一部分比较适合用来训练一个有效的分类模型,并且适应测试数据。于是,目标就是从源域数据中找出那些适合测试数据的实例,并将这些实例迁移到用于目标域数据的模型的学习中去。在基于实例的迁移学习方面,包括AdaBoost算法,以及具有迁移能力的boosting算法Tradaboosting,使之具有迁移学习的能力,从而能够最大限度的

利用源域来帮助目标的分类其实现方式是建立一种自动调整权重的机制,重要的源域数据权重将会增加,不重要的源域数据的权重将会减小。基于数据迭代分类过程中权重动态更新,从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据。

[0041] 除了选取出训练目标域数据之外,还需要从目标域数据选取出训练目标域数据。选取的原则是选取出适合训练模型使用的有效样本数据。例如从数据完整性(数据是否包含标签)等方面进行选取。

[0042] S203:利用训练数据集进行风险模型训练。

[0043] 在一种可选方式中,利用训练数据集进行风险分类模型训练包括:针对训练数据集,通过二分类算法(如逻辑回归算法、决策树算法、梯度提升树算法或随机森林算法)进行训练,得到风险识别模型。

[0044] 在一种可选方式中,在完成模型建立之后,执行对交易数据的风险识别和控制:根据风险识别模型,对业务数据进行识别,确定出业务风险分值;对于风险分值高于风险阈值的业务进行控制(例如拦截交易)。

[0045] 下面以一个具体应用场景对本说明书实施例提供的基于迁移学习的风控方法进行示例性描述。

[0046] 该应用场景中涉及不同国家的数据,利用tradaboost算法实现在不同国家之间的风控模型迁移的问题。例如该风控场景中涉及多个国家,包括菲律宾(PH)、马来西亚(MY)、新加坡(SG)等国家,但每个国家之间的交易量、案件浓度等分布是不同的,具体黑样本浓度为PH>>SG>>MY。业务目标是实现一个风险控制模型能否有效识别出各个国家之间的风险交易,常规的做法是所有国家样本数据一起进行建模,但问题是这样的模型建设出来在PH国家上面识别的性能很好,而在MY、SG两个国家之间性能相对差很多,原因就是它们两国之间黑样本浓度太少,整体交易流量占所有国家也很少。因此,引入迁移学习的思路,实现通过从PH的样本数据上学习到有利于对MY(或SG)进行建模的样本,从而提升在MY(或SG)上的模型性能,在本例子中使用tradaboost算法。在本例子中,target domain指代MY(或SG),source domain是PH。

[0047] 参见图3为该实例实现示意图,涉及四个步骤:

[0048] 第一步、将source domain和target domain的数据进行合并,它们具有相同的数据结构,具体来说最后的数据宽表应该有相同的变量维度,每个变量有着相同的业务逻辑定义。

[0049] 第二步、将合并的数据利用tradaboost算法找出输出有利于解决目标问题的样本数据(后续将详细描述如何tradaboost算法确定样本数据)。

[0050] 第三步、基于第二步输出的样本数据,利用传统的二分类算法(比如逻辑回归、决策树、GBDT、随机森林等算法)进行建模。

[0051] 第四步、根据模型进行交易数据的风险识别和控制。

[0052] 下面介绍基于tradaboost算法实现从source domain中找出有利于解决target domain问题的样本数据,来补充target domain本身样本数据不足的问题,从而提升模型性能。

[0053] 参见图4,为本实例中tradaboost算法逻辑示意图。tradaboost是利用AdaBoost算法的思想原理,起初给训练数据(由target domain和source domain数据汇总而成)中的每

一个样例都赋予一个权重,接下来每一轮迭代过程中,当一个target domain中的样本被错误的分类之后,认为这个样本是很难分类的,于是可以加大这个样本的权重,这样在下一次的训练中这个样本所占的比重就更大了。如果source domain中的一个样本被错误的分类了,认为这个样本对于目标数据是很不同的,就降低这个数据在样本中所占的权重,降低这个样本在分类器中所占的比重。整个过程就是从source domain中找出有利于解决目标问题的样本数据,来补充target domain的样本数据。

[0054] 可见在此实例中,从源域数据中选取出训练源域数据的具体实现过程是:为源域数据中每一个样本数据赋予权重;在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则降低样本数据的权重;根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从源域数据中选取出与目标域数据具有类似或相同数据分布的训练源域数据。相应的,从目标域数据选取出训练目标域数据具体实现过程是:为目标域数据中每一个样本数据赋予权重;在迭代分类过程中,判断样本数据是否被错误分类,若被错误分类,则提高样本数据的权重;根据权重高低优先选取权重高的样本数据进行下一次迭代分类,从而从目标域数据选取出训练目标域数据。

[0055] 在本说明书实施例提供的基于迁移学习的风控方法中,基于迁移学习算法,从源域数据中选取出训练源域数据作为目标域数据的有效补充,在目标域数据不足等情况下,有助于模型建立的准确性。通过打通不同数据源(例如不同国家数据或不同市场数据),提升风控效率。

[0056] 第二方面,基于同一发明构思,本说明书实施例提供一种基于迁移学习的风控装置,请参考图5,包括:

[0057] 数据获取单元501,用于获取风险交易样本的源域数据及目标域数据;

[0058] 训练数据集确定单元502,用于基于迁移学习算法,从所述源域数据中选取出与所述目标域数据具有类似或相同数据分布的训练源域数据,并从所述目标域数据选取出训练目标域数据;合并所述训练源域数据与所述训练目标域数据为训练数据集;

[0059] 训练单元503,用于利用所述训练数据集进行风险模型训练。

[0060] 在一种可选方式中,还包括:

[0061] 初始合并单元504,用于合并所述源域数据及目标域数据为具有相同数据结构的初始合并数据集;其中,所述训练数据集是从所述初始合并数据集中选定的。

[0062] 在一种可选方式中,所述训练数据集确定单元502具体用于:

[0063] 基于数据特征提取及特征聚类,从所述源域数据中选取出与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

[0064] 基于数据深度学习建模及模型参数分布相似度,从所述源域数据中选取出与所述目标域数据具有类似或相同数据分布的训练源域数据;或者,

[0065] 基于数据迭代分类过程中权重动态更新,从所述源域数据中选取出与所述目标域数据具有类似或相同数据分布的训练源域数据。

[0066] 在一种可选方式中,所述训练数据集确定单元502具体用于:根据所述目标域数据中各数据的完整性,选取出训练目标域数据。

[0067] 在一种可选方式中,所述训练数据集确定单元502包括:

[0068] 源域权重赋予子单元,用于为所述源域数据中每一个样本数据赋予权重;

[0069] 源域权重动态更新子单元，用于在迭代分类过程中，判断样本数据是否被错误分类，若被错误分类，则降低所述样本数据的权重；

[0070] 源域数据确定子单元，用于根据权重高低优先选取权重高的样本数据进行下一次迭代分类，从而从所述源域数据中选取与所述目标域数据具有类似或相同数据分布的训练源域数据。

[0071] 在一种可选方式中，所述训练数据集确定单元502还包括：

[0072] 目标域权重赋予子单元，用于为所述目标域数据中每一个样本数据赋予权重；

[0073] 目标域权重动态更新子单元，用于在迭代分类过程中，判断样本数据是否被错误分类，若被错误分类，则提高所述样本数据的权重；

[0074] 目标域数据确定子单元，用于根据权重高低优先选取权重高的样本数据进行下一次迭代分类，从而从所述目标域数据选取训练目标域数据。

[0075] 在一种可选方式中，所述训练单元503具体用于：针对所述训练数据集，通过二分类算法进行训练，得到风险识别模型。

[0076] 在一种可选方式中，还包括：

[0077] 风险识别单元505，用于根据所述风险识别模型，对业务数据进行识别，确定出业务风险分值；

[0078] 风险控制单元506，用于对于风险分值高于风险阈值的业务进行控制。

[0079] 第三方面，基于与前述实施例中基于迁移学习的风控方法同样的发明构思，本发明还提供一种服务器，如图6所示，包括存储器604、处理器602及存储在存储器604上并可在处理器602上运行的计算机程序，所述处理器602执行所述程序时实现前文所述基于迁移学习的风控方法的任一方法的步骤。

[0080] 其中，在图6中，总线架构(用总线600来代表)，总线600可以包括任意数量的互联的总线和桥，总线600将包括由处理器602代表的一个或多个处理器和存储器604代表的存储器的各种电路链接在一起。总线600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606在总线600和接收器601和发送器603之间提供接口。接收器601和发送器603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602负责管理总线600和通常的处理，而存储器604可以被用于存储处理器602在执行操作时所使用的数据。

[0081] 第四方面，基于与前述实施例中基于迁移学习的风控方法的发明构思，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基于迁移学习的风控方法的任一方法的步骤。

[0082] 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。

[0083] 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。

[0084] 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。

[0085] 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。

[0086] 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。

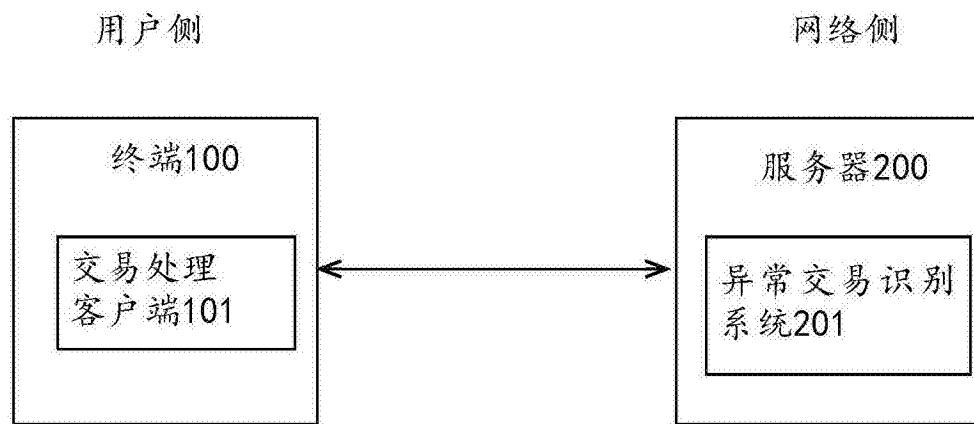


图1

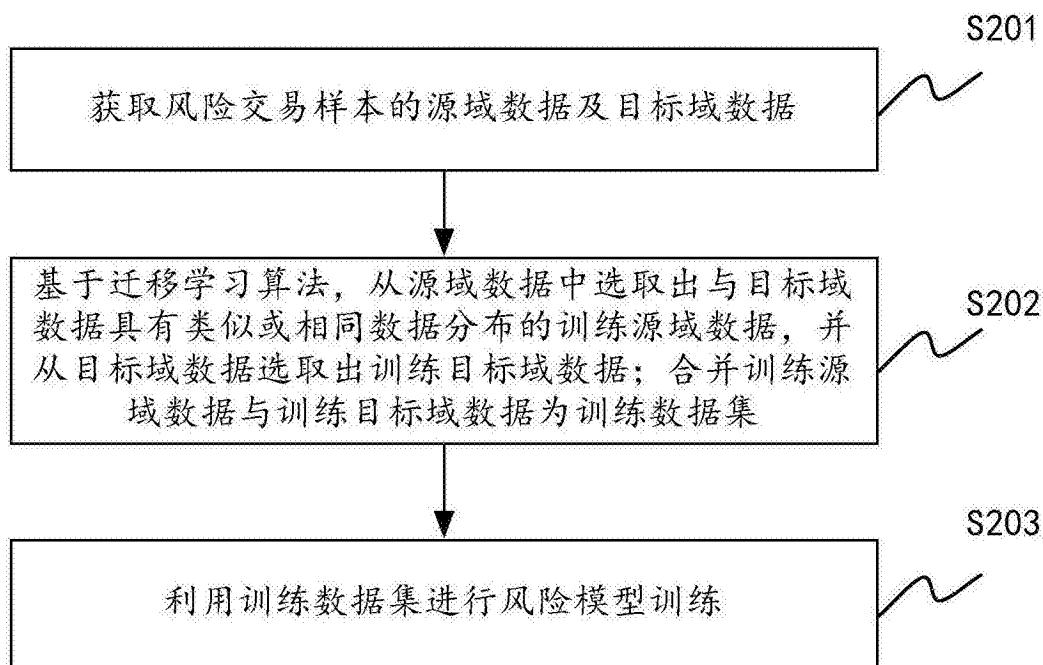


图2

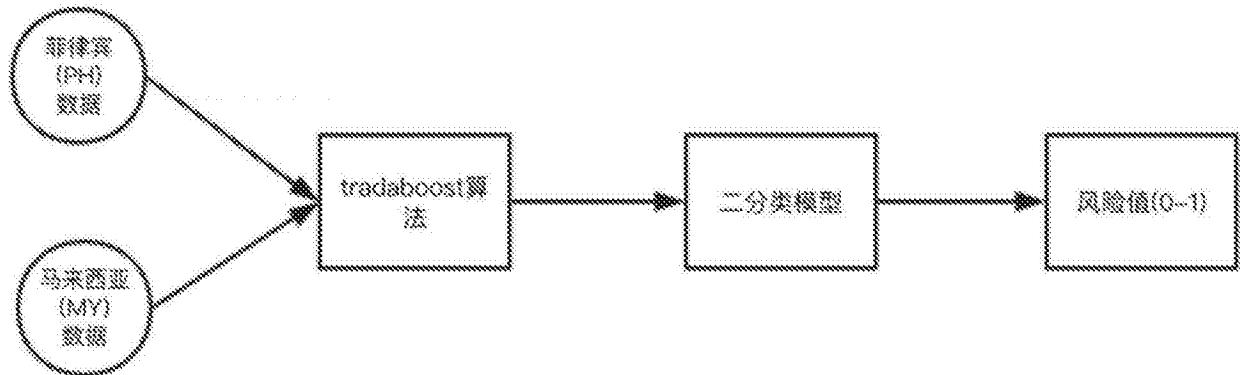


图3

降低错误分类的数据权重      增加错误分类的数据权重

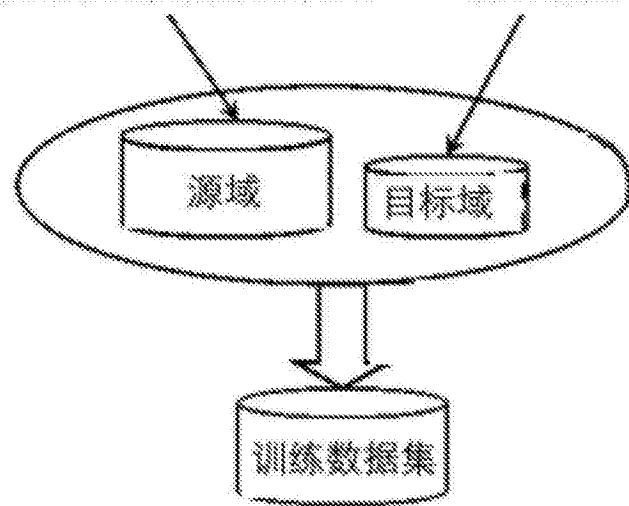


图4

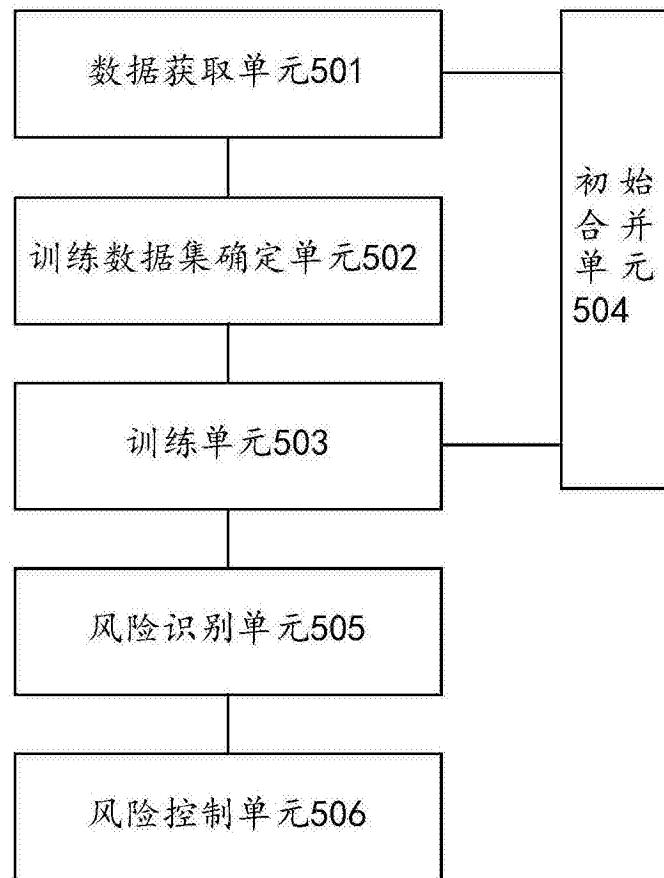


图5

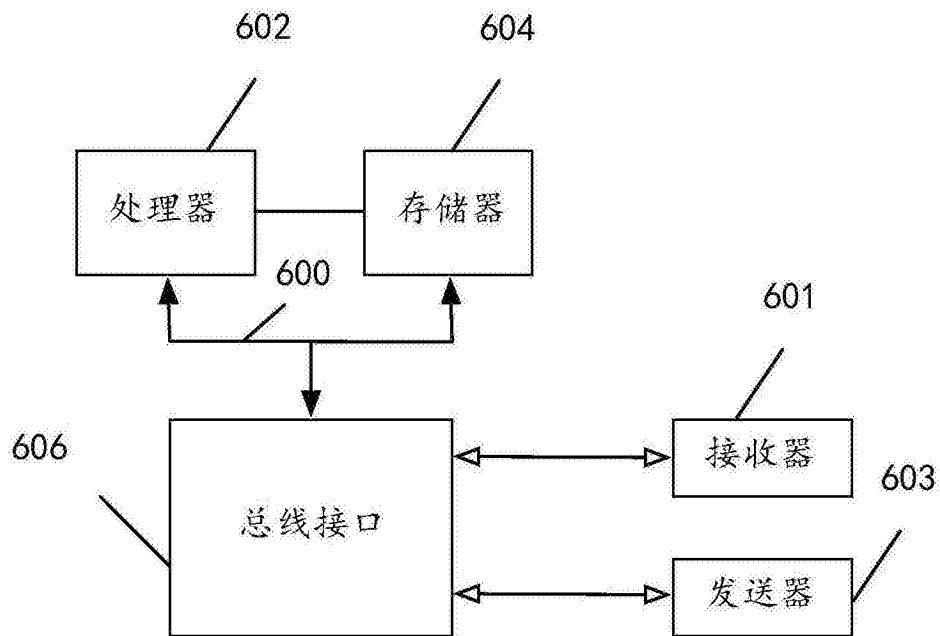


图6