



(12)发明专利申请

(10)申请公布号 CN 113824580 A

(43)申请公布日 2021.12.21

(21)申请号 202010568192.1

(22)申请日 2020.06.19

(71)申请人 北京市天元网络技术股份有限公司

地址 100193 北京市海淀区西北旺东路10号院东区20号楼浪潮大厦

(72)发明人 王晓晗 苏鹏

(74)专利代理机构 北京路浩知识产权代理有限公司 11002

代理人 郑朝然

(51)Int.Cl.

H04L 12/24(2006.01)

G06N 20/10(2019.01)

G06K 9/62(2006.01)

权利要求书2页 说明书7页 附图2页

(54)发明名称

一种网络指标预警方法及系统

(57)摘要

本发明实施例提供一种网络指标预警方法及系统。该方法包括：获取待预测网络数据集合；将所述待预测网络数据集合输入至预先训练好的告警预测模型中，得到所述告警预测模型输出的告警预测结果；其中，所述告警预测模型是基于网络指标样本集数据，按照网络告警标准样板进行分类所得的分级标签进行训练，并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。本发明实施例通过采用多个机器学习算法对网络性能指标进行训练，并采用交叉验证和集成学习获得高精度预测模型，该模型具有自动学习，使得网络告警预测得到迅速而准确的响应，使得维护人员快速识别有效告警信息。



1. 一种网络指标预警方法,其特征在于,包括:

获取待预测网络数据集合;

将所述待预测网络数据集合输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

2. 根据权利要求1所述的网络指标预警方法,其特征在于,所述获取待预测网络数据集合,具体包括:

在预设时间段内以预设时间间隔收集若干预设网络性能指标;

将所述若干预设网络性能指标作为所述待预测网络数据集合。

3. 根据权利要求1或2所述的网络指标预警方法,其特征在于,所述告警预测模型,通过以下步骤获得:

根据预设样本收集周期获取所述网络指标样本集数据,并对所述网络指标样本集数据进行预处理,得到预处理样本集数据;

按照网络告警标准样板对所述预处理样本集数据进行分类,得到所述分级标签;

将标注对应分级标签的所述预处理样本集数据存放于若干个文件夹,生成用于训练的若干个数据集;

获取若干机器学习算法分别作为初始模型集合;

将所述若干个数据集输入所述初始模型集合中进行训练,得到初始训练模型集合;

使用所述交叉验证法验证所述初始训练模型集合的输出精度,并根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型。

4. 根据权利要求3所述的网络指标预警方法,其特征在于,所述根据预设样本收集周期获取所述网络指标样本集数据,并对所述网络指标样本集数据进行预处理,得到预处理样本集数据,具体包括:

按照网络预测精度要求调整所述预设样本收集周期,在所述预设样本收集周期内收集网络性能指标数据集合;

将所述网络性能指标数据集合进行数据清理,并按照预设格式进行输出,得到所述预处理样本集数据。

5. 根据权利要求3所述的网络指标预警方法,其特征在于,所述若干机器学习算法包括逻辑回归、XGboost、GBDT、随机森林和SVM。

6. 根据权利要求3所述的网络指标预警方法,其特征在于,所述使用所述交叉验证法验证所述初始训练模型集合的输出精度,具体包括:

将所述若干个数据集随机分为K份集合,其中K-1份作为训练集,剩下一份作为测试集;

重复训练K次,得到K个模型和K个性能评估结果;

基于所述K个性能评估结果,计算平均性能评估结果;

在所述若干个数据集上重新训练模型,用所述K份集合中独立的测试集进行测试,得到最终性能评估结果,将所述最终性能评估结果作为所述初始训练模型集合的输出精度。

7. 根据权利要求3所述的网络指标预警方法,其特征在于,所述根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型,具体包括:

改变所述若干个数据集的结构,得到若干变化的数据集;
将所述若干变化的数据集按照不同的线性组合构造成若干分类器;
根据若干预设评价指标选取所述若干分类器中精度最高的模型作为所述告警预测模型。

8. 一种网络指标预警系统,其特征在于,包括:

获取模块,用于获取待预测网络数据集;

处理模块,用于将所述待预测网络数据集输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

9. 一种电子设备,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,其特征在于,所述处理器执行所述程序时实现如权利要求1至7任一项所述网络指标预警方法的步骤。

10. 一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在于,该计算机程序被处理器执行时实现如权利要求1至7任一项所述网络指标预警方法的步骤。

一种网络指标预警方法及系统

技术领域

[0001] 本发明涉及网络技术领域,尤其涉及一种网络指标预警方法及系统。

背景技术

[0002] 随着技术的发展,网络结构越来越复杂,入网人数增多导致了网络负载压力增大。

[0003] 在运营商的网络维护场景中,运营商的告警阈值一般需要人工配置,配置过程复杂,学习门槛高,需要耗费大量人力与时间;告警缺乏实时性,在告警后发生较长时间后才能预警,延长了维保人员抢修时间。

[0004] 因此,如何提高网络预警的效率和准确性,成为网络维护中亟待解决的问题。

发明内容

[0005] 本发明实施例提供一种网络指标预警方法及系统,用以解决现有技术中网络告警需要人工配置,导致精度不高,效率低下的缺陷。

[0006] 第一方面,本发明实施例提供一种网络指标预警方法,包括:

[0007] 获取待预测网络数据集合;

[0008] 将所述待预测网络数据集合输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0009] 进一步地,所述获取待预测网络数据集合,具体包括:

[0010] 在预设时间段内以预设时间间隔收集若干预设网络性能指标;

[0011] 将所述若干预设网络性能指标作为所述待预测网络数据集合。

[0012] 进一步地,所述告警预测模型,通过以下步骤获得:

[0013] 根据预设样本收集周期获取所述网络指标样本集数据,并对所述网络指标样本集数据进行预处理,得到预处理样本集数据;

[0014] 按照网络告警标准样板对所述预处理样本集数据进行分类,得到所述分级标签;

[0015] 将标注对应分级标签的所述预处理样本集数据存放于若干个文件夹,生成用于训练的若干个数据集;

[0016] 获取若干机器学习算法分别作为初始模型集合;

[0017] 将所述若干个数据集输入所述初始模型集合中进行训练,得到初始训练模型集合;

[0018] 使用所述交叉验证法验证所述初始训练模型集合的输出精度,并根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型。

[0019] 进一步地,所述根据预设样本收集周期获取所述网络指标样本集数据,并对所述网络指标样本集数据进行预处理,得到预处理样本集数据,具体包括:

[0020] 按照网络预测精度要求调整所述预设样本收集周期,在所述预设样本收集周期内

收集网络性能指标数据集合；

[0021] 将所述网络性能指标数据集合进行数据清理,并按照预设格式进行输出,得到所述预处理样本集数据。

[0022] 进一步地,所述若干机器学习算法包括逻辑回归、XGboost、GBDT、随机森林和SVM。

[0023] 进一步地,所述使用所述交叉验证法验证所述初始训练模型集合的输出精度,具体包括:

[0024] 将所述若干个数据集随机分为K份集合,其中K-1份作为训练集,剩下一份作为测试集;

[0025] 重复训练K次,得到K个模型和K个性能评估结果;

[0026] 基于所述K个性能评估结果,计算平均性能评估结果;

[0027] 在所述若干个数据集上重新训练模型,用所述K份集合中独立的测试集进行测试,得到最终性能评估结果,将所述最终性能评估结果作为所述初始训练模型集合的输出精度。

[0028] 进一步地,所述根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型,具体包括:

[0029] 改变所述若干个数据集的结构,得到若干变化的数据集;

[0030] 将所述若干变化的数据集按照不同的线性组合构造成若干分类器;

[0031] 根据若干预设评价指标选取所述若干分类器中精度最高的模型作为所述告警预测模型。

[0032] 第二方面,本发明实施例提供一种网络指标预警系统,包括:

[0033] 获取模块,用于获取待预测网络数据集合;

[0034] 处理模块,用于将所述待预测网络数据输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0035] 第三方面,本发明实施例提供一种电子设备,包括:

[0036] 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任一项所述网络指标预警方法的步骤。

[0037] 第四方面,本发明实施例提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现任一项所述网络指标预警方法的步骤。

[0038] 本发明实施例提供的网络指标预警方法及系统,通过采用多个机器学习算法对网络性能指标进行训练,并采用交叉验证和集成学习获得高精度预测模型,该模型具有自动学习,使得网络告警预测得到迅速而准确的响应,并得维护人员快速识别有效告警信息。

附图说明

[0039] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术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。

- [0040] 图1为本发明实施例提供的一种网络指标预警方法流程图；
- [0041] 图2为本发明实施例提供的交叉验证示意图；
- [0042] 图3为本发明实施例提供的一种网络指标预警系统结构图；
- [0043] 图4为本发明实施例提供的电子设备的结构框图。

具体实施方式

[0044] 为使本发明实施例的目的、技术方案和优点更加清楚，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。

[0045] 针对现有技术中针对网络维护告警信息存在的问题，本发明实施例引入机器学习算法，将运营商网络中收集到的性能数据在指定目录下按时间顺序汇总，作为模型训练的输入，并由训练好的模型输出预测结果，其中应用了交叉验证对多个机器学习算法的训练结果进行验证，并结合集成学习法进行模型选优，得到预测精度较高的告警预测模型。

[0046] 图1为本发明实施例提供的一种网络指标预警方法流程图，如图1所示，包括：

[0047] S1，获取待预测网络数据集合；

[0048] S2，所述待预测网络数据集合输入至预先训练好的告警预测模型中，得到所述告警预测模型输出的告警预测结果；其中，所述告警预测模型是基于网络指标样本集数据，按照网络告警标准样板进行分类所得的分级标签进行训练，并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0049] 具体地，将一系列网络运行的指标数据输入到已经训练好的告警预测模型中，便得到下一个预设时刻内是否存在告警信息，该训练好的告警预测模型是由搜集的大量网络性能数据构成的样本集数据，按照分级标签进行告警等级划分，并采用交叉验证法对经过多个机器学习算法训练的结果进行精度验证，以及还通过集成学习法进行优选模型选择而产生的。

[0050] 本发明实施例通过多维度的进行模型训练的交叉验证和集成学习的优选输出，使得网络告警预测得到迅速而准确的响应，并得维护人员快速识别有效告警信息。

[0051] 基于上述实施例，所述获取待预测网络数据集合，具体包括：

[0052] 在预设时间段内以预设时间间隔收集若干预设网络性能指标；

[0053] 将所述若干预设网络性能指标作为所述待预测网络数据集合。

[0054] 具体地，根据实际网络运行的需求和测试精度的要求，可动态地调整模型的数据输入，使得模型较为灵活地适应动态网络的变化，例如按照上一个小时的预设时间段，周期性地收集若干预设的网络性能指标，如LU成功率数据以及告警信息，间隔周期可设为例如5分钟，并将这些大量的数据作为待预测网络数据集合。

[0055] 本发明实施例通过灵活调整模型的输入数据范围，增强了模型的动态调整和泛化能力。

[0056] 基于上述任一实施例，所述告警预测模型，通过以下步骤获得：

[0057] 根据预设样本收集周期获取所述网络指标样本集数据，并对所述网络指标样本集数据进行预处理，得到预处理样本集数据；

- [0058] 按照网络告警标准样板对所述预处理样本集数据进行分类,得到所述分级标签;
- [0059] 将标注对应分级标签的所述预处理样本集数据存放于若干个文件夹,生成用于训练的若干个数据集;
- [0060] 获取若干机器学习算法分别作为初始模型集合;
- [0061] 将所述若干个数据集输入所述初始模型集合中进行训练,得到初始训练模型集合;
- [0062] 使用所述交叉验证法验证所述初始训练模型集合的输出精度,并根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型。
- [0063] 其中,所述根据预设样本收集周期获取所述网络指标样本集数据,并对所述网络指标样本集数据进行预处理,得到预处理样本集数据,具体包括:
- [0064] 按照网络预测精度要求调整所述预设样本收集周期,在所述预设样本收集周期内收集网络性能指标数据集;
- [0065] 将所述网络性能指标数据集进行数据清理,并按照预设格式进行输出,得到所述预处理样本集数据。
- [0066] 其中,所述若干机器学习算法包括逻辑回归、XGboost、GBDT、随机森林和SVM。
- [0067] 其中,所述使用所述交叉验证法验证所述初始训练模型集合的输出精度,具体包括:
- [0068] 将所述若干个数据集随机分为K份集合,其中K-1份作为训练集,剩下一份作为测试集;
- [0069] 重复训练K次,得到K个模型和K个性能评估结果;
- [0070] 基于所述K个性能评估结果,计算平均性能评估结果;
- [0071] 在所述若干个数据集上重新训练模型,用所述K份集合中独立的测试集进行测试,得到最终性能评估结果,将所述最终性能评估结果作为所述初始训练模型集合的输出精度。
- [0072] 其中,所述根据所述集成学习法选取所述初始训练模型集合中精度最高的模型作为所述告警预测模型,具体包括:
- [0073] 改变所述若干个数据集的结构,得到若干变化的数据集;
- [0074] 将所述若干变化的数据集按照不同的线性组合构造成若干分类器;
- [0075] 根据若干预设评价指标选取所述若干分类器中精度最高的模型作为所述告警预测模型。
- [0076] 具体地,首先搜集用于模型训练的样本集数据,收集的范围是动态变化的,例如按照每月/每季度/每年,防止模型不能更好的适应变化的数据,再进行数据清理和按照规定格式进行处理和输出存储的数据,得到预处理样本集数据。此处,根据告警对应的级别判定将对应的预处理样本集数据进行分类,得到模型训练用的分级标签,对应放入多个文件夹,生成多个训练用的数据集。
- [0077] 再采用多个机器学习算法作为初始网络模型结构,本发明实施例采用如下几种机器学习算法:
- [0078] (1) 逻辑回归:又称逻辑回归分析,是一种广义的线性回归分析模型,常用于数据挖掘,疾病自动诊断,经济预测等领域;

[0079] (2) XGBoost:是一个优化的分布式梯度增强库,它在Gradient Boosting框架下实现机器学习算法,XGBoost提供并行树提升(也称为GBDT,GBM),可以快速准确地解决许多数据科学问题;

[0080] (3) GBDT:全称为梯度下降树,通过采用加法模型(即基函数的线性组合),以及不断减小训练过程产生的残差来达到将数据分类或者回归的算法;

[0081] (4) 随机森林:指的是利用多棵树对样本进行训练并预测的一种分类器,其输出的类别是由个别树输出的类别的众数而定;

[0082] (5) SVM:支持向量机(Support Vector Machine,SVM)是一类按监督学习(supervised learning)方式对数据进行二元分类的广义线性分类器(generalized linear classifier),其决策边界是对学习样本求解的最大边距超平面(maximum-margin hyperplane)。

[0083] 采用上述几种机器学习算法得到多个初始训练结果后,再采用交叉验证的方式来测试模型的泛化能力,交叉验证法示意图如图2所示,使用训练数据集对训练模型进行评价,与Handout法不同无需新数据测试,具体流程为:

[0084] 将数据集无替换的随机分为K份,k-1份用来训练模型,剩下的一份用来性能评估;

[0085] 重复k次,得到K个模型和性能评估结果;

[0086] 在得到K个性能评估后,计算平均性能评估;

[0087] 在整个数据集上重新训练模型,再用独立的测试集得到最终的性能评估。

[0088] 交叉验证法对一个数据集用同一个模型进行K次不同的测试,但每个训练的数据集由于随机抽取又不完全一样,相当于扩充了数据集;若这K个模型的均值效果较好,说明训练的模型有一定泛化能力。

[0089] 最后使用集成学习法(Ensemble Learning),也称提升(Boosting)方法,广泛用于分类和回归任务,主要思路为:使用一些(不同的)方法改变原始训练样本的分布,从而构建多个不同的分类器,并将这些分类器线性组合得到一个更强大的分类器,来做最后的决策。

[0090] 此处,当前主流集成学习方法有三种:

[0091] 1.Boosting方法:包括Adaboosting,提升树(代表是GBDT),XGBoost等;

[0092] 2.Bagging方法:典型的是随机森林;

[0093] 3.Stacking算法。

[0094] 得到告警预测模型后,进行结果的多维度验证,表1为综合了在五种机器学习算法下的交叉验证正确率和从正确率、查准率和查全率三个维度进行集成学习验证反应的对新数据的预测情况。

模型名称	交叉验证正确率	对新数据的预测		
		正确率	查准率	查全率
[0095] 逻辑回归	0.7356	0.5405	0.6461	0.1791
XGboost	0.7630	0.5544	0.6584	0.2260
GBDT	0.7509	0.5597	0.6505	0.2580
随机森林	0.7147	0.7217	0.7587	0.6503
SVM	0.6599	0.5853	0.7898	0.2324

[0096] 该阶段的目的是将所有模型的结果进行对比,查看出模型的泛化能力,通过表中数据可知:当前随机森林算法学习得到的模型对新数据的预测时成功率、查准率和查全率都明显高于其他模型,泛化性能最好,可以根据集成学习法选为当前模型的算法。

[0097] 图3为本发明实施例提供的一种网络指标预警系统结构图,如图3所示,包括:获取模块31和处理模块32;其中:

[0098] 获取模块31用于获取待预测网络数据集合;处理模块32用于将所述待预测网络数据集合输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0099] 本发明实施例提供的系统用于执行上述对应的方法,其具体的实施方式与方法实施方式一致,涉及的算法流程与对应的方法算法流程相同,此处不再赘述。

[0100] 本发明实施例通过多维度的进行模型训练的交叉验证和集成学习的优选输出,使得网络告警预测得到迅速而准确的响应,并得维护人员快速识别有效告警信息。

[0101] 图4示例了一种电子设备的实体结构示意图,如图4所示,该电子设备可以包括:处理器(processor)410、通信接口(Communications Interface)420、存储器(memory)430和通信总线440,其中,处理器410,通信接口420,存储器430通过通信总线440完成相互间的通信。处理器410可以调用存储器430中的逻辑指令,以执行如下方法:获取待预测网络数据集合;将所述待预测网络数据集合输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0102] 此外,上述的存储器430中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用,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,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施

例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0103] 另一方面,本发明实施例还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各实施例提供的传输方法,例如包括:获取待预测网络数据集合;将所述待预测网络数据集合输入至预先训练好的告警预测模型中,得到所述告警预测模型输出的告警预测结果;其中,所述告警预测模型是基于网络指标样本集数据,按照网络告警标准样板进行分类所得的分级标签进行训练,并通过交叉验证法获取模型输出精度以及通过集成学习法获取若干机器学习方法中最优输出模型所得到的。

[0104] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0105] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行各个实施例或者实施例的某些部分所述的方法。

[0106] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

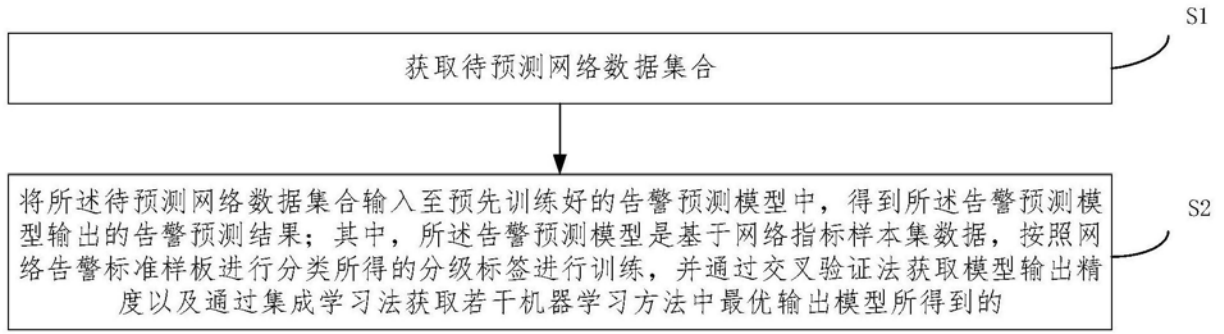


图1



图2

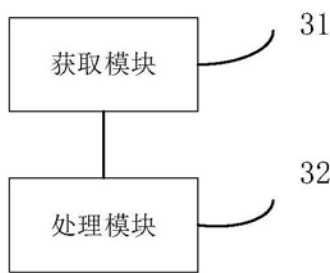


图3

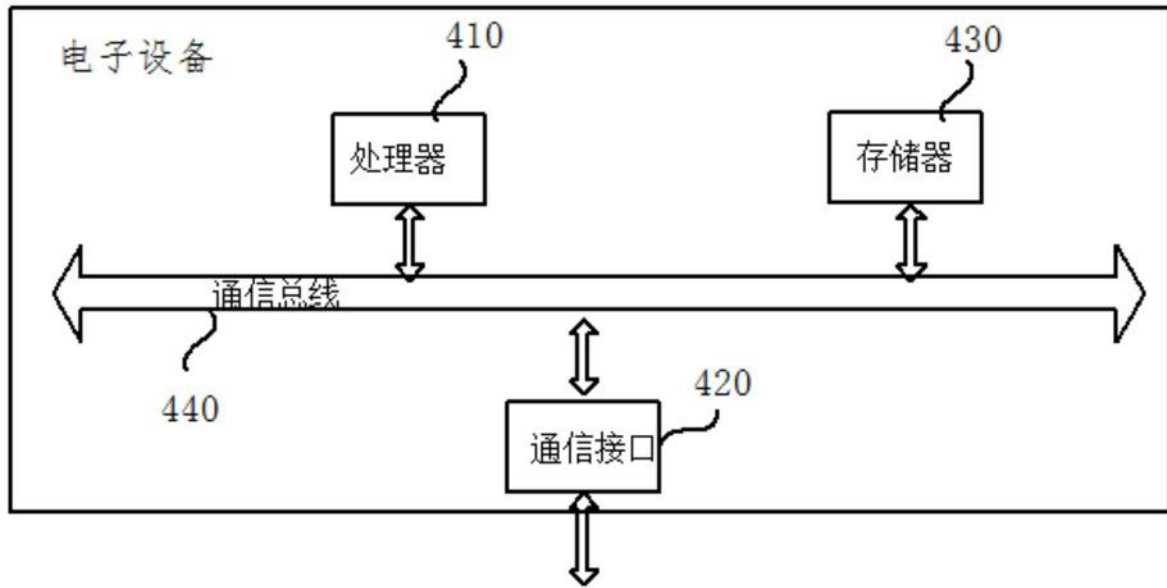


图4