



(12)发明专利申请

(10)申请公布号 CN 113452802 A

(43)申请公布日 2021.09.28

(21)申请号 202010211207.9

(22)申请日 2020.03.24

(71)申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 王薇 谭涵秋 钟培楠

(74)专利代理机构 北京三高永信知识产权代理有限责任公司 11138

代理人 颜晶

(51)Int.Cl.

H04L 29/12(2006.01)

G06K 9/62(2006.01)

G06F 16/2455(2019.01)

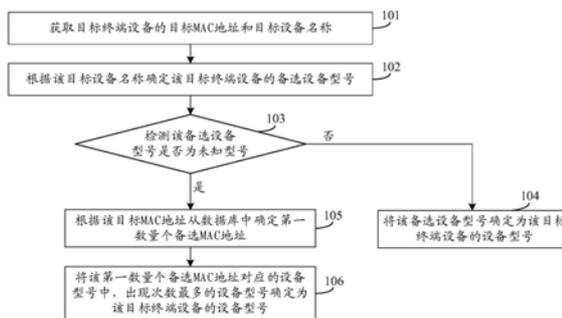
权利要求书4页 说明书22页 附图7页

(54)发明名称

设备型号的识别方法、装置及系统

(57)摘要

本申请提供了一种设备型号的识别方法、装置及系统,属于数据处理技术领域。该方法可以根据获取到的目标物理地址,先从数据库中确定出与该目标物理地址相似度较高的第一数量个备选物理地址,然后将该第一数量个备选物理地址对应的设备型号中,出现次数最多的设备型号确定为目标终端设备的设备型号。由此,即使数据库中未存储该目标物理地址对应的设备型号,也可以根据物理地址的相似度确定出该目标终端设备的设备型号,从而有效提高了设备型号识别的成功率,降低了对数据库中存储的数据量的要求。



1. 一种设备型号的识别方法,其特征在于,所述方法包括:

获取目标终端设备的目标物理地址;

从数据库中确定第一数量个备选物理地址,所述数据库包括多个物理地址,以及每个所述物理地址对应的设备型号,其中每个所述备选物理地址与所述目标物理地址的相似度,大于所述数据库中其他物理地址与所述目标物理地址的相似度,所述第一数量为大于1的整数;

将所述第一数量个备选物理地址对应的设备型号中,出现次数最多的设备型号确定为所述目标终端设备的设备型号。

2. 根据权利要求1所述的方法,其特征在于,所述数据库包括多个数据组,每个所述数据组包括一个或多个数据对,每个所述数据对包括一个物理地址,以及与所述物理地址对应的设备型号;所述从数据库中确定第一数量个备选物理地址,包括:

根据所述目标物理地址与每个所述数据组中任一数据对的物理地址的相似度,确定第二数量个备选数据组,每个所述备选数据组中所述任一数据对的物理地址与所述目标物理地址的相似度,大于其他数据组中所述任一数据对的物理地址与所述目标物理地址的相似度,其中,所述第二数量为大于1的整数;

从所述第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址,每个所述备选物理地址与所述目标物理地址的相似度,大于所述第二数量个备选数据组中的其他物理地址与所述目标物理地址的相似度。

3. 根据权利要求2所述的方法,其特征在于,每个所述数据组中存在一个中心数据对,且每个所述数据组中的数据对按照与所述中心数据对的物理地址的相似度由高到低的顺序排列;每个所述数据组中的任一数据对为所述中心数据对;

所述从所述第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址,包括:

根据每个所述备选数据组中的数据对的排列顺序,从所述第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址。

4. 根据权利要求3所述的方法,其特征在于,所述方法还包括:

获取多个所述数据对;

采用聚类算法对多个所述数据对进行分组,得到多个所述数据组;

对于每个所述数据组,根据所述数据组中的中心数据对的物理地址与其他每个数据对中的物理地址的相似度,对所述数据组包括的数据对按照相似度由高到低的顺序进行排序。

5. 根据权利要求2所述的方法,其特征在于,在所述从数据库中确定第一数量个备选物理地址之前,所述方法还包括:

采用相似度模型确定所述数据库中的物理地址与所述目标物理地址的相似度;

其中,所述相似度模型基于已确定相似度的多个物理地址样本训练得到。

6. 根据权利要求1至5任一所述的方法,其特征在于,所述方法还包括:

获取所述目标终端设备的目标设备名称;

根据所述目标设备名称确定所述目标终端设备的备选设备型号;

若所述备选设备型号不为未知型号,则将所述备选设备型号确定为所述目标终端设备

的设备型号；

所述从数据库中确定第一数量个备选物理地址，包括：

若所述备选设备型号为所述未知型号，则从数据库中确定第一数量个备选物理地址。

7. 根据权利要求6所述的方法，其特征在于，所述根据所述目标设备名称确定所述目标终端设备的备选设备型号，包括：

采用型号确定模型从所述目标设备名称中确定所述目标终端设备的备选设备型号；

其中，所述型号确定模型基于已确定设备型号的多个设备名称样本训练得到。

8. 根据权利要求7所述的方法，其特征在于，所述采用型号确定模型从所述目标设备名称中确定所述目标终端设备的备选设备型号，包括：

采用型号确定模型确定所述目标设备名称中的每个字符是否为有效字符；

将所述目标设备名称中的有效字符组成的字符串确定为所述目标终端设备的备选设备型号。

9. 根据权利要求8所述的方法，其特征在于，所述方法还包括：

获取设备名称样本以及所述设备名称样本对应的设备型号样本；

将所述设备名称样本中，与所述设备型号样本匹配的字符串中的每个字符均标注为有效字符，将除所述字符串之外的其他字符均标注为无效字符；

对标注后的所述设备名称样本进行模型训练，得到所述型号确定模型。

10. 根据权利要求6所述的方法，其特征在于，所述根据所述目标设备名称确定所述目标终端设备的备选设备型号，包括：

分别确定所述目标设备名称与多个设备型号模板中每个设备型号模板的匹配度；

将匹配度最高的设备型号模板确定为所述目标终端设备的备选设备型号。

11. 一种设备型号的识别装置，其特征在于，所述装置包括：

第一获取模块，用于获取目标终端设备的目标物理地址；

第一确定模块，用于从数据库中确定第一数量个备选物理地址，所述数据库包括多个物理地址，以及每个所述物理地址对应的设备型号，其中每个所述备选物理地址与所述目标物理地址的相似度，大于所述数据库中其他物理地址与所述目标物理地址的相似度，所述第一数量为大于1的整数；

第二确定模块，用于将所述第一数量个备选物理地址对应的设备型号中，出现次数最多的设备型号确定为所述目标终端设备的设备型号。

12. 根据权利要求11所述的装置，其特征在于，所述数据库包括多个数据组，每个所述数据组包括一个或多个数据对，每个所述数据对包括一个物理地址，以及与所述物理地址对应的设备型号；所述第一确定模块，用于：

根据所述目标物理地址与每个所述数据组中任一数据对的物理地址的相似度，确定第二数量个备选数据组，每个所述备选数据组中所述任一数据对的物理地址与所述目标物理地址的相似度，大于其他数据组中所述任一数据对的物理地址与所述目标物理地址的相似度，其中，所述第二数量为大于1的整数；

从所述第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址，每个所述备选物理地址与所述目标物理地址的相似度，大于所述第二数量个备选数据组中的其他物理地址与所述目标物理地址的相似度。

13. 根据权利要求12所述的装置,其特征在于,每个所述数据组中存在一个中心数据对,且每个所述数据组中的数据对按照与所述中心数据对的物理地址的相似度由高到低的顺序排列;每个所述数据组中的任一数据对为所述中心数据对;

所述第一确定模块,用于:根据每个所述备选数据组中的数据对的排列顺序,从所述第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址。

14. 根据权利要求13所述的装置,其特征在于,所述装置还包括:

第二获取模块,用于获取多个所述数据对;

聚类模块,用于采用聚类算法对多个所述数据对进行分组,得到多个所述数据组;

排序模块,用于对于每个所述数据组,根据所述数据组中的中心数据对的物理地址与其他每个数据对中的物理地址的相似度,对所述数据组包括的数据对按照相似度由高到低的顺序进行排序。

15. 根据权利要求12所述的装置,其特征在于,所述装置还包括:

第三确定模块,用于在所述从数据库中确定第一数量个备选物理地址之前,采用相似度模型确定所述数据库中的物理地址与所述目标物理地址的相似度;

其中,所述相似度模型基于已确定相似度的多个物理地址样本训练得到。

16. 根据权利要求11至15任一所述的装置,其特征在于,所述装置还包括:

所述第一获取模块,还用于获取所述目标终端设备的目标设备名称;

第四确定模块,用于根据所述目标设备名称确定所述目标终端设备的备选设备型号;

第五确定模块,用于若所述备选设备型号不为未知型号,则将所述备选设备型号确定为所述目标终端设备的设备型号;

所述第一确定模块,用于若所述备选设备型号为所述未知型号,则从数据库中确定第一数量个备选物理地址。

17. 根据权利要求16所述的装置,其特征在于,所述第四确定模块,用于:

采用型号确定模型从所述目标设备名称中确定所述目标终端设备的备选设备型号;

其中,所述型号确定模型基于已确定设备型号的多个设备名称样本训练得到。

18. 根据权利要求17所述的装置,其特征在于,所述第四确定模块,用于:

采用型号确定模型确定所述目标设备名称中的每个字符是否为有效字符;

将所述目标设备名称中的有效字符组成的字符串确定为所述目标终端设备的备选设备型号。

19. 根据权利要求18所述的装置,其特征在于,所述装置还包括:

第三获取模块,用于获取设备名称样本以及所述设备名称样本对应的设备型号样本;

第六确定模块,用于将所述设备名称样本中,与所述设备型号样本匹配的字符串中的每个字符均标注为有效字符,将除所述字符串之外的其他字符均标注为无效字符;

训练模块,用于对标注后的所述设备名称样本进行模型训练,得到所述型号确定模型。

20. 根据权利要求16所述的装置,其特征在于,所述第四确定模块,用于:

分别确定所述目标设备名称与多个设备型号模板中每个设备型号模板的匹配度;

将匹配度最高的设备型号模板确定为所述目标终端设备的备选设备型号。

21. 一种设备型号的识别装置,其特征在于,所述装置包括:处理器,存储器,以及存储在所述存储器上并能够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

序时实现如权利要求1至10任一所述的方法。

22. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有指令,当所述计算机可读存储介质在计算机上运行时,使得计算机执行如权利要求1至10任一所述的方法。

23. 一种设备型号的识别系统,其特征在于,所述系统包括:第一服务器和第二服务器;所述第一服务器用于执行如权利要求1至3,权利要求5至8以及权利要求10中任一所述的方法;

所述第二服务器用于执行如权利要求4或9所述的方法。

24. 一种设备型号的识别系统,其特征在于,所述系统包括:第一服务器,所述第一服务器用于执行如权利要求1至10任一所述的方法。

25. 根据权利要求23或24所述的系统,其特征在于,所述系统还包括:网关设备;

所述网关设备分别与终端设备和所述第一服务器连接,所述网关设备用于获取所述终端设备的物理地址,并将获取到的所述物理地址发送至所述第一服务器。

## 设备型号的认识方法、装置及系统

### 技术领域

[0001] 本申请涉及数据处理领域,特别涉及一种设备型号的认识方法、装置及系统。

### 背景技术

[0002] 在对网络中的终端设备(例如手机、路由器或智能设备等)进行维护或故障检测时,可以通过该终端设备的设备型号获取该终端设备的规格参数,进而可以基于该规格参数进行较为高效的维护或故障检测。

[0003] 相关技术中,网关(gateway)设备可以获取其所连接的终端设备的介质访问控制(media access control,MAC)地址(也称为物理地址),并可以通过查询数据库中预先存储的MAC地址与设备型号的对应关系,确定终端设备的设备型号。

[0004] 但是,若数据库中未记录终端设备的MAC地址,则无法确定终端设备的设备型号,相关技术中的方法对数据库中存储的数据量要求较高,设备型号认识的成功率较低。

### 发明内容

[0005] 本申请提供了一种设备型号的认识方法、装置及系统,可以解决相关技术中的方法对数据库存储的数据量要求较高,认识成功率较低的问题,技术方案如下:

[0006] 一方面,提供了一种设备型号的认识方法,该方法可以获取目标终端设备的目标MAC地址;并从数据库中确定第一数量个备选MAC地址,该数据库包括多个MAC地址,以及每个MAC地址对应的设备型号,其中每个备选MAC地址与该目标MAC地址的相似度,大于该数据库中其他MAC地址与该目标MAC地址的相似度,该第一数量为大于1的整数;之后可以将该第一数量个备选MAC地址对应的设备型号中,出现次数最多的设备型号确定为该目标终端设备的设备型号。

[0007] 基于本申请提供的方案,即使数据库中未存储目标MAC地址对应的设备型号,也可以根据MAC地址的相似度确定出目标终端设备的设备型号,从而有效提高了设备型号认识的成功率,降低了对数据库中存储的数据量的要求。

[0008] 可选的,该数据库可以包括多个数据组,每个数据组包括一个或多个数据对,每个数据对包括一个MAC地址,以及与该MAC地址对应的设备型号;从该数据库中确定第一数量个备选MAC地址的过程可以包括:

[0009] 根据该目标MAC地址与每个数据组中任一数据对的MAC地址的相似度,确定第二数量个备选数据组,每个备选数据组中该任一数据对的MAC地址与该目标MAC地址的相似度,大于其他数据组中该任一数据对的MAC地址与该目标MAC地址的相似度,其中,该第二数量为大于1的整数;从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址,每个备选MAC地址与该目标MAC地址的相似度,大于该第二数量个备选数据组中的其他MAC地址与该目标MAC地址的相似度。

[0010] 通过先确定第二数量个备选数据组,可以缩小MAC地址的搜索范围,从而有效提高备选MAC地址的搜索效率。

[0011] 可选的,每个数据组中存在一个中心数据对,且每个数据组中的数据对按照与该中心数据对的MAC地址的相似度由高到低的顺序排列;每个数据组中的任一数据对为该中心数据对;相应的,从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址,包括:根据每个备选数据组中的数据对的排列顺序,从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址。

[0012] 由于每个数据组中的数据对按照相似度由高到低的顺序排列,因此在采用近邻搜索算法确定备选MAC地址时,能够较为快速的从备选数据组中确定出与目标MAC地址近邻(即相似度较高)的备选MAC地址,有效提高了备选MAC地址的搜索效率。

[0013] 可选的,该方法还可以包括:获取多个数据对;采用聚类算法对多个数据对进行分组,得到多个数据组;对于每个数据组,根据该数据组中的中心数据对的MAC地址与其他每个数据对中的MAC地址的相似度,对该数据组包括的数据对按照相似度由高到低的顺序进行排序。采用上述方法构建的数据库可以便于搜索与目标MAC地址近邻的备选MAC地址。

[0014] 可选的,在该从数据库中确定第一数量个备选MAC地址之前,该方法还可以包括:采用相似度模型确定该数据库中的MAC地址与该目标MAC地址的相似度;其中,该相似度模型基于已确定相似度的多个MAC地址样本训练得到。

[0015] 该相似度模型可以基于深度度量学习算法训练得到,采用该相似度模型确定出的两个MAC地址的相似度可以较为准确的反映该两个MAC地址对应的设备型号的相似程度。

[0016] 可选的,该方法还可以包括:获取该目标终端设备的目标设备名称;根据该目标设备名称确定该目标终端设备的备选设备型号;若该备选设备型号不为未知型号,即根据该目标设备名称能够确定出该目标终端设备的设备型号,则将该备选设备型号确定为该目标终端设备的设备型号;相应的,从数据库中确定第一数量个备选MAC地址的过程可以包括:若该备选设备型号为未知型号,即根据该目标设备名称无法确定出该目标终端设备的设备型号,则从数据库中确定第一数量个备选MAC地址。

[0017] 本申请提供的方案还可以采用目标设备名称确定设备型号,相比于仅基于单一参数确定设备型号,有效提高了设备型号的识别成功率。

[0018] 可选的,根据该目标设备名称确定该目标终端设备的备选设备型号的过程可以包括:采用型号确定模型从该目标设备名称中确定该目标终端设备的备选设备型号;其中,该型号确定模型基于已确定设备型号的多个设备名称样本训练得到。

[0019] 可选的,采用型号确定模型从该目标设备名称中确定该目标终端设备的备选设备型号的过程可以包括:采用型号确定模型确定该目标设备名称中的每个字符是否为有效字符;将该目标设备名称中的有效字符组成的字符串确定为目标终端设备的备选设备型号。

[0020] 采用神经网络模型从该目标设备名称中确定设备型号,可以确保设备型号的识别成功率,以及确保确定出的设备型号的可靠性。

[0021] 可选的,该方法还可以包括:获取设备名称样本以及该设备名称样本对应的设备型号样本;将该设备名称样本中,与该设备型号样本匹配的字符串中的每个字符均标注为有效字符,将除该字符串之外的其他字符均标注为无效字符;对标注后的设备名称样本进行模型训练,得到该型号确定模型。

[0022] 可选的,根据该目标设备名称确定该目标终端设备的备选设备型号的过程可以包括:

[0023] 分别确定该目标设备名称与多个设备型号模板中每个设备型号模板的匹配度;将匹配度最高的设备型号模板确定为该目标终端设备的备选设备型号。

[0024] 基于模板匹配的方法无需预先训练神经网络模型,该方法的复杂度较低。

[0025] 另一方面,提供了一种设备型号的识别装置,该装置可以包括至少一个模块,该至少一个模块可以用于实现上述方面所提供的设备型号的识别方法。

[0026] 又一方面,提供了一种设备型号的识别装置,该装置可以包括:处理器,存储器,以及存储在该存储器上并能够在该处理器上运行的计算机程序,该处理器执行该计算机程序时实现如上述方面所提供的设备型号的识别方法。

[0027] 再一方面,提供了一种计算机可读存储介质,该计算机可读存储介质中存储有指令,当该计算机可读存储介质在计算机上运行时,可以使得计算机执行如上述方面所提供的设备型号的识别方法。

[0028] 再一方面,提供了一种设备型号的识别系统,该系统可以包括:第一服务器和第二服务器;该第一服务器可以用于执行上述方面所提供的设备型号的识别方法中确定设备型号的步骤;该第二服务器可以用于执行上述方面所提供的设备型号的识别方法中模型训练和/或数据库构建的步骤。

[0029] 再一方面,提供了一种设备型号的识别系统,该系统可以包括:第一服务器,该第一服务器可以用于执行如上述方面所提供的设备型号的识别方法。

[0030] 可选的,该系统还可以包括:网关设备;该网关设备分别与终端设备和该第一服务器连接,该网关设备用于获取该终端设备的MAC地址,并将获取到的该MAC地址发送至该第一服务器。

[0031] 综上所述,本申请提供了一种设备型号的识别方法、装置及系统,本申请提供的方案可以根据目标MAC地址与数据库中存储的MAC地址的相似度,从数据库中确定出相似度较高的第一数量个备选MAC地址,然后再将该第一数量个备选MAC地址对应的设备型号中,出现次数最多的设备型号确定为目标终端设备的设备型号。由此,即使数据库中未存储该目标MAC地址对应的设备型号,也可以根据MAC地址的相似度确定出该目标终端设备的设备型号,从而有效提高了设备型号识别的成功率,降低了对数据库中存储的数据量的要求。

## 附图说明

[0032] 图1是本申请实施例提供的一种设备型号的识别系统的结构示意图;

[0033] 图2是本申请实施例提供的一种设备型号的识别方法的流程图;

[0034] 图3是本申请实施例提供的另一种设备型号的识别方法的流程图;

[0035] 图4是本申请实施例提供的一种确定目标终端设备的备选设备型号的方法流程图;

[0036] 图5是本申请实施例提供的一种目标设备名称的识向量的示意图;

[0037] 图6是本申请实施例提供的一种确定第一数量个备选MAC地址的方法流程图;

[0038] 图7是本申请实施例提供的一种数据库的结构示意图;

[0039] 图8是本申请实施例提供的一种采用型号识别模型确定设备型号的示意图;

[0040] 图9是本申请实施例提供的一种型号确定模型的训练方法的流程图;

[0041] 图10是本申请实施例提供的一种相似度模型的训练方法的流程图;

- [0042] 图11是本申请实施例提供的一种数据库的构建方法的流程图；
- [0043] 图12是本申请实施例提供的一种划分数据组的示意图；
- [0044] 图13是本申请实施例提供的一种VP-tree的结构示意图；
- [0045] 图14是本申请实施例提供的一种设备型号的识别装置的结构示意图；
- [0046] 图15是本申请实施例提供的另一种设备型号的识别装置的结构示意图；
- [0047] 图16是本申请实施例提供的又一种设备型号的识别装置的结构示意图；
- [0048] 图17是本申请实施例提供的再一种设备型号的识别装置的结构示意图。

## 具体实施方式

[0049] 下面结合附图详细介绍本申请实施例提供的设备型号的识别方法、装置及系统。

[0050] 终端设备的设备型号识别对于运营商的家庭网络业务来说有着重要的意义。众所周知,在家庭网络中决定用户体验的关键是无线保真(wireless fidelity,WiFi)的网速,而现网数据分析表明,导致WiFi网速较差的原因中,约有80%是因为手机和路由器等终端设备自身性能不支持相应WiFi带宽,而WiFi信号穿墙等因素只占20%。譬如,很多网关设备支持2.4吉赫兹(GHz)和5GHz的频段,而家庭所用的路由器或者手机有可能只支持2.4GHz的频段,由此即可导致用户不能享受5G频段带来的高速体验,进而导致用户投诉。

[0051] 目前,运营商在处理家庭网络业务的质量投诉时通常采取上门服务,导致人工运维成本过高。而通过识别网关设备所连接的终端设备的设备型号,确定其对WiFi频段的支持能力,不仅可以快速定位家庭网络质量差的根本因素,从而降低运维人员上门服务的次数,亦可以提前分析家庭对于WiFi的需求,从而推荐合适的网络套餐。总而言之,设备型号的识别既可以降低运营商的人工运维成本,又可以提升用户的宽带使用体验。

[0052] 图1是本申请实施例提供的一种设备型号的识别系统的结构示意图,如图1所示,该系统可以包括型号识别服务器01。其中,该型号识别服务器01可以是一台服务器,也可以是由若干台服务器组成的服务器集群,或者还可以是一个云计算服务中心。该系统还可以包括网关设备02,该网关设备02用于实现不同网络互连的设备,又可以称为网间连接器或协议转换器。该系统还可以包括网管设备03,该网管设备03可以是一台电脑。

[0053] 参考图1可以看出,该网关设备02可以连接有多个终端设备04,每个终端设备04可以为手机、电脑、路由器、可穿戴设备或家居设备等智能终端设备。其中家居设备可以包括音箱、电子秤、电视和空调等。

[0054] 图2是本申请实施例提供的一种设备型号的识别方法的流程图,结合图1和图2,该网关设备02可以用于实现步骤A:数据采集,即该网关设备02可以采集终端设备04的MAC地址(或如图2所示的MAC地址和设备名称),并将采集到的数据发送至型号识别服务器01。例如参考图2,网关设备02采集并上报的数据可以包括设备名称N1至N3,以及每个设备名称对应的MAC地址,即MAC地址M1至M3。

[0055] 型号识别服务器01可以用于实现步骤B:设备型号的识别。例如参考图2,该型号识别服务器01可以根据MAC地址识别设备型号(步骤B2);该型号识别服务器01也可以根据设备名称识别设备型号(步骤B1);或者,该型号识别服务器01可以执行步骤B2和步骤B1,并且合并通过步骤B1和步骤B2得到的识别结果,得到最终的设备型号。例如图2所示,型号识别服务器01基于设备名称N1和MAC地址M1,最终识别出的设备型号可以为X1。该型号识别服务

器01进而还可以将最终确定出的设备型号发送至网管设备03。

[0056] 该网管设备03可以显示型号识别服务器01识别出的设备型号(步骤C)。管理该网管设备03的网络管理员(例如运营商的运维人员)进而可以基于识别出的设备型号确定终端设备04的规格参数,以便进行网络维护、故障检测或需求分析。

[0057] 可选的,如图1所示,该型号识别服务器01可以包括第一服务器011以及第二服务器012。该第二服务器012可以用于构建数据库,以及根据标签数据(labeled data)训练用于进行设备型号识别的神经网络模型,并将构建的数据库以及训练好的神经网络模型发送至第一服务器011。其中,标签数据是指已确定设备型号的样本数据。第一服务器011进而可以基于该数据库和神经网络模型识别终端设备04的设备型号。因此,其中第一服务器011也可以称为在线识别服务器,该第二服务器012也可以称为离线训练服务器。

[0058] 在本申请实施例中,该网络管理员还可以对网管设备03提供的终端设备04的设备型号进行校验。若确定该设备型号有误,则网络管理员可以在网管设备03中输入正确的设备型号,并指示网管设备03将纠正数据(包括正确的设备型号、MAC地址和设备名称)发送至第二服务器012。第二服务器012可以周期性地基于累积的纠正数据对神经网络模型进行重训练,得到更新后的神经网络模型,并将该更新后的神经网络模型发送至该第一服务器011。也即是,该网管设备03还可以指示第二服务器012实现图2所示的步骤D。

[0059] 需要说明的是,该第一服务器011和第二服务器012可以为集成的一个服务器,也可以为相互独立的两个服务器。或者,该第一服务器011与该网管设备03可以为集成的一个设备。又或者,该系统中也可以无需网关设备02,终端设备04可以直接向型号识别服务器01上报MAC地址,或者该网管设备03收集相关数据并向型号识别服务器01上报。

[0060] 本申请实施例提供了一种设备型号的识别方法,该方法可以应用于如图1所示的型号识别服务器01中,例如可以应用于该第一服务器011中。参考图3,该方法可以包括:

[0061] 步骤101、获取目标终端设备的目标MAC地址和目标设备名称。

[0062] 该目标终端设备是指待查询设备型号的终端设备。

[0063] 作为一种可选的实现方式,型号识别服务器01可以获取网关设备02上报的目标终端设备的目标MAC地址和目标设备名称。

[0064] 参考图1,网关设备02可以周期性采集其所连接的终端设备的MAC地址和设备名称,并向型号识别服务器01上报终端设备的MAC地址和设备名称。其中,该网关设备02采集MAC地址和设备名称的周期可以根据需求灵活调整,例如可以为1小时、一天或一个月等。

[0065] 或者,网关设备02可以响应于型号识别服务器01或网管设备03发送的采集指令,采集目标终端设备的目标MAC地址和目标设备名称并上报至型号识别服务器01。其中,该采集指令中可以携带有能够唯一标识目标终端设备的标识,例如互联网协议(Internet protocol, IP)地址。

[0066] 作为另一种可选的实现方式,型号识别服务器01也可以直接获取目标终端设备的目标MAC地址和目标设备名称。

[0067] 示例的,如图1所示,假设该网关设备02连接有手机、电脑、电视和路由器共4个终端设备04,则该网关设备02采集并上报的该4个终端设备04的MAC地址和设备名称可以如表1所示。参考表1,网管设备02可以向型号识别服务器01上报4个设备信息,每个设备信息也可以称为一条设备记录(record)。其中每个设备信息可以包括MAC地址和设备名称两个字

段。例如,该型号识别服务器01获取到的网管设备02上报的手机的设备信息可以包括目标MAC地址:001234AB56C1,以及目标设备名称:小明的AA-a1-666。

[0068] 表1

| MAC地址        | 设备名称         |
|--------------|--------------|
| 001234AB56C1 | 小明的AA-a1-666 |
| 001234EF56B2 | 小明的电脑        |
| 005678AB56C1 | 小明家的电视       |
| 008912AC56C1 | CC-c1        |

[0070] 步骤102、根据该目标设备名称确定该目标终端设备的备选设备型号。

[0071] 在本申请实施例中,型号识别服务器01获取到目标终端设备的目标MAC地址和目标设备名称后,可以先根据该目标设备名称确定该目标终端设备的备选设备型号。

[0072] 作为一种可选的实现方式,该型号识别服务器01中可以存储有型号确定模型,该型号确定模型可以是基于自然语言处理(natural language processing,NLP)技术,采用已确定设备型号的多个设备名称样本训练得到的。型号识别服务器01获取到目标终端设备的目标设备名称后,可以采用该型号确定模型从该目标设备名称中确定该目标终端设备的备选设备型号。

[0073] 如图4所示,采用该型号确定模型从该目标设备名称中确定该目标终端设备的备选设备型号的过程可以包括:

[0074] 步骤1021、对目标设备名称进行编码,得到机器可识别的名称向量。

[0075] 由于设备名称通常是由汉字、数字、字母以及标点符号等字符组成,而这些字符中的大部分字符(例如汉字、字母和标点符号等)是机器学习算法不能理解的,因此型号识别服务器01需要先对目标设备名称进行编码,得到机器可识别的名称向量。

[0076] 可选的,可以采用独热(one-hot)编码算法对目标设备名称进行编码,从而将目标设备名称中的每个字符均转化成机器可以识别的字符,这些机器可识别的字符组成了名称向量。其中,采用one-hot编码算法对目标设备名称进行编码的过程相当于根据预先创建的映射表,将目标设备名称中的每个字符映射成一个整数。

[0077] 例如,a-z这26个阿拉伯字母,可以被映射成0-25这26个整数;0-9这10个数字,可以被映射成26-35这些整数;常在设备名称中出现的标点符号,诸如“-”,可以被映射成35-38这3个整数;常用的汉字也可以被映射成38之后的整数。譬如“小”可以被映射成39,“明”可以被映射成40,“的”可以被映射成41。

[0078] 示例的,假设型号识别服务器01获取到的目标设备名称为“小明的AA-a1-666”,则该型号识别服务器01对该目标设备名称进行one-hot编码后,可以得到名称向量[39,40,41,1,1,36,1,27,36,32,32,32],该名称向量为矢量。

[0079] 步骤1022、将名称向量输入至型号确定模型,得到型号确定模型输出的标识向量。

[0080] 型号识别服务器01可以将编码得到的名称向量输入至型号确定模型,得到型号确定模型输出的标识向量。该标识向量可以包括多个标识符,每个标识符可以用于标识该目标设备名称中的一个字符是否为有效字符。也即是,型号识别服务器01可以采用该型号确定模型确定该目标设备名称中的每个字符是否为有效字符。其中有效字符是指用于组成目标终端设备的设备型号的字符。在本申请实施例中,可以通过标识符的不同取值表示其所

指示的字符是否为有效字符。例如,标识符的取值为第一整数时可以表示其所指示的字符为有效字符,标识符的取值为第二整数时可以表示其所指示的字符为无效字符。该第一整数与该第二整数不同。

[0081] 示例的,该标识向量可以是一个长度与该名称向量的长度相等的整数序列,且每个标识符的取值范围可以为[0,1,2]。其中,标识符取值为0可以表示其指示的字符为有效字符,且该字符为设备型号的起始字符(即首个字符)。标识符取值为1可以表示其指示的字符为有效字符,且该字符为设备型号的中间字符。标识符取值为2则可以表示其指示的字符为无效字符。也即是,第一整数包括0和1,第二整数为2。

[0082] 参考图5,假设目标终端设备的设备型号为AA-a1,目标设备名称为“小明的AA-a1-666”,则型号识别服务器01将名称向量[39,40,41,1,1,36,1,27,36,32,32,32]输入至型号确定模型后,该型号确定模型输出的标识向量可以为[2,2,2,0,1,1,1,1,1,2,2,2]。

[0083] 步骤1023、根据该标识向量从目标设备名称中确定目标终端设备的备选设备型号。

[0084] 型号识别服务器01可以根据该标识向量,将该目标设备名称中的有效字符组成的字符串确定为该目标终端设备的备选设备型号。

[0085] 示例的,如图5所示,型号识别服务器01可以将标识向量[2,2,2,0,1,1,1,1,1,2,2,2]中,标识符0和标识符1指示的字符串“AA-a1”确定为目标终端设备的备选设备型号。

[0086] 需要说明的是,对于某些无意义的目标设备名称,型号确定模型输出的标识向量中的每个标识符指示的字符可能均为无效字符,对于目标设备名称中的字符均为无效字符的情况,型号识别服务器01可以确定该目标终端设备的设备型号未知(unknown),即确定该目标终端设备的备选设备型号为未知型号。比如对于目标设备名称“小明的电脑”,型号确定模型输出的标识向量可以是[2,2,2,2,2],则型号识别服务器01基于该标识向量可以确定该目标终端设备的备选设备型号为unknown。

[0087] 作为另一种可选的实现方式,该型号识别服务器01中可以预先存储有多个设备型号模板。型号识别服务器01获取到目标终端设备的目标设备名称后,可以分别确定该目标设备名称与每个设备型号模板的匹配度,并可以将匹配度最高的设备型号模板确定为该目标终端设备的备选设备型号。

[0088] 可选的,该型号识别服务器01可以采用最大公约子串匹配法,计算该目标设备名称与每个设备型号模板的最大公约子串长度,并可以将长度最大的设备型号模板确定为该目标终端设备的备选设备型号。

[0089] 示例的,假设型号识别服务器01获取到的目标设备名称为“小明的AA-a1-666”,且型号识别服务器01中存储的设备型号模板包括“AA-a1”、“AA-a2”以及“BB-b1”。则型号识别服务器01采用最大公约子串匹配法进行匹配度计算后,可以确定与该目标设备名称为“小明的AA-a1-666”匹配度最高的设备型号模板为“AA-a1”,因此可以将该设备型号模板“AA-a1”确定为该目标终端设备的备选设备型号。

[0090] 基于模板匹配的方法无需预先训练神经网络模型,该方法的复杂度较低。

[0091] 步骤103、检测该备选设备型号是否为未知型号。

[0092] 由于用户可以自主设置终端设备的设备名称,因此型号识别服务器01基于目标设备名称并不一定能确定出可靠的备选设备型号。例如现网数据表明,大约55%的设备名称

能够能给出有效的设备型号的信息,而剩余约45%的设备名称则是无效的。

[0093] 因此在本申请实施例中,型号识别服务器01在基于目标设备名称确定出备选设备型号之后,还可以进一步检测该备选设备型号是否为未知型号。若该备选设备型号不为未知型号,则型号识别服务器01可以执行步骤104;若该备选设备型号为未知型号,则型号识别服务器01可以执行步骤105。

[0094] 示例的,型号识别服务器01可以检测该备选设备型号是否为unknown,若不为unknown,则可以执行步骤104;若为unknown,则可以执行步骤105。

[0095] 步骤104、将该备选设备型号确定为该目标终端设备的设备型号。

[0096] 若基于目标设备名称确定出的备选设备型号不为未知型号,则型号识别服务器01可以直接将该备选设备型号确定为该目标终端设备的设备型号。

[0097] 步骤105、根据该目标MAC地址从数据库中确定第一数量个备选MAC地址。

[0098] 若基于目标设备名称确定出的备选设备型号为未知型号,则型号识别服务器01可以继续根据该目标终端设备的目标MAC地址确定目标终端设备的设备型号,从而提升该设备型号识别的成功率。

[0099] 在本申请实施例中,型号识别服务器01中可以存储有数据库,该数据库包括多个MAC地址,以及每个MAC地址对应的设备型号。型号识别服务器01获取到目标MAC地址后,可以计算该目标MAC地址与该数据库中每个MAC地址的相似度,并可以基于计算得到的相似度,从数据库中确定出第一数量个备选MAC地址。

[0100] 其中,每个备选MAC地址与该目标MAC地址的相似度,大于该数据库中其他MAC地址与该目标MAC地址的相似度。也即是,该型号识别服务器01可以从数据库中确定出与该目标MAC地址相似度最高的第一数量个备选MAC地址。该第一数量可以为大于等于1的整数,例如可以为5或者10等。

[0101] 可选的,型号识别服务器01可以采用相似度模型确定该数据库中的MAC地址与该目标MAC地址的相似度。该相似度模型可以是预先基于已确定相似度的多个MAC地址样本训练得到的。采用该相似度模型确定出的两个MAC地址的相似度,可以反映该两个MAC地址对应的设备型号相同的概率。即两个MAC地址的相似度越高,表明该两个MAC地址对应的设备型号相同的概率越高。

[0102] 在本申请实施例中,为了提高查询备选MAC地址的效率,该数据库可以采用局部敏感哈希(locality-sensitive hashing,LSH)技术构建得到。采用该LSH技术构建的数据库可以包括多个数据组,每个数据组包括一个或多个数据对,每个数据对包括一个MAC地址,以及与该MAC地址对应的设备型号。

[0103] 其中,该多个数据组可以是采用聚类算法划分得到的,因此每个数据组也可以称为一个数据簇(cluster)。每个数据对在数据库中的存储位置可以用两级索引(索引1,索引2)表示。其中,索引1(index1)为数据组的索引,索引2(index2)为数据对在数据组中的索引。由于基于该LSH技术构建的数据库可以将设备型号相同或相似的数据对划分至同一个数据组,因此设备型号相同或相似的数据对的索引1相同。

[0104] 表2

[0105]

| 索引1/索引2 | 001  | 002  | 003  |
|---------|------|------|------|
| 001     | 数据对1 | 数据对2 | 数据对3 |

|     |      |      |      |
|-----|------|------|------|
| 002 | 数据对4 | 数据对5 | 数据对6 |
| 003 | 数据对7 | 数据对8 | 数据对9 |

[0106] 示例的,参考表2,假设该数据库中存储有9个数据对,该9个数据对被划分为3个数据组,该3个数据组的索引分别为001、002和003。则每个数据组中各个数据对中的设备型号可以均相同。例如,索引1为001的3个数据对中的设备型号可以均为AA-a1。

[0107] 图6是本申请实施例提供的一种确定第一数量个备选MAC地址的方法流程图,如图6所示,上述步骤105可以包括:

[0108] 步骤1051、根据目标MAC地址与每个数据组中任一数据对的MAC地址的相似度,确定第二数量个备选数据组。

[0109] 在本申请实施例中,由于数据库中的多个数据对已经基于聚类算法被划分为了多个数据组,每个数据组中各个数据对的MAC地址的相似度较高,而属于不同数据组的两个数据对的MAC地址的相似度较低。因此为了提高查询的效率,型号识别服务器01可以先计算该目标MAC地址与每个数据组中任一数据对的MAC地址的相似度,得到与该数据库包括的数据组的个数相同数量的相似度。之后,型号识别服务器01即可基于计算得到的相似度,确定第二数量个备选数据组。

[0110] 其中,每个备选数据组中该任一数据对的MAC地址与该目标MAC地址的相似度,大于其他数据组中该任一数据对的MAC地址与该目标MAC地址的相似度。也即是,该型号识别服务器01可以先确定出与该目标MAC地址最相似的第二数量个备选数据组。该第二数量可以为大于等于1的整数,且该第二数量与该第一数量可以相同,也可以不同。并且,各个数据组中所选取用于与该目标MAC地址进行相似度计算的任一数据对的索引2可以相同。例如,每个数据组中所选取的任一数据对的索引2可以均为001,即可以分别计算该目标MAC地址与每个数据组中的第一个数据对的MAC地址的相似度。

[0111] 可选的,该数据库中的多个数据组可以是基于k中心(k-centers)聚类算法划分得到的,则每个数据组中存在一个中心数据对。并且,每个数据组中的中心数据对可以位于第一位,其他数据对可以按照与该中心数据对的MAC地址的相似度由高到低的顺序排列。相应的,型号识别服务器01在计算该目标MAC地址与每个数据组中任一数据对的MAC地址的相似度时,可以选择每个数据组中的中心数据对进行计算。

[0112] 示例的,参考表2,假设索引1为001的数据组中的中心数据对为数据对1,索引1为002的数据组中的中心数据对为数据对4,索引1为003的数据组中的中心数据对为数据对7,则该3个中心数据对的索引2可以均为001,即均可以位于其所属的数据组中的第一位。型号识别服务器01在获取到目标MAC地址后,可以分别计算该目标MAC地址与数据对1中的MAC地址的相似度 $y_1$ ,该目标MAC地址与数据对4中的MAC地址的相似度 $y_2$ ,以及该目标MAC地址与数据对7中的MAC地址的相似度 $y_3$ 。若该第二数量为2,且型号识别服务器01计算得到三个相似度满足: $y_1 > y_3 > y_2$ ,则型号识别服务器01可以索引1为001的数据组以及索引1为003的数据组确定为备选数据组。

[0113] 步骤1052、从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址。

[0114] 对于任意一个备选数据组,其中的每个备选MAC地址与该目标MAC地址的相似度,可以大于该备选数据组中的其他MAC地址与该目标MAC地址的相似度。也即是,该型号识别

服务器01可以从该第二数量个备选数据组包括的MAC地址中确定出与该目标MAC地址最相似的第一数量个备选MAC地址。

[0115] 在本申请实施例中,为了提高备选MAC地址的查询效率,该型号识别服务器01可以采用近邻搜索算法,从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址。

[0116] 例如,若每个数据组中的数据对均按照与该中心数据对的MAC地址的相似度由高到低的顺序排列,则型号识别服务器01可以根据每个备选数据组中的数据对的排列顺序,从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址。

[0117] 可选的,每个数据组包括的一个或多个数据对可以采用制高点树(vantage point tree,VP-tree)算法进行排序。相应的,型号识别服务器01可以采用该VP-tree算法从该第二数量个备选数据组包括的MAC地址中确定出第一数量个备选MAC地址,由此可以大大加速每个数据组内的数据对的近邻搜索效率。

[0118] 示例的,假设该第一数量为10,第二数量为5,则型号识别服务器01计算目标MAC地址与每个数据组中的中心数据对(即索引2为001的数据对)的MAC地址的相似度后,可以从数据库中确定出5个备选数据组。假设该5个备选数据组分别为c1,c2,c3,c4和c5,则得益于数据库构建中采取的聚类算法,与目标MAC地址最相似的10个MAC地址较大概率在c1至c5这5个备选数据组中。

[0119] 之后,对于c1至c5中的每个备选数据组,型号识别服务器01可以根据VP-tree查询该备选数据组中物理地址与该目标MAC地址最相似的若干个备选数据对。其中,不同备选数据组中确定出的备选数据对的个数可以相同,也可以不同。假设型号识别服务器01从每个备选数据组中均确定出5个备选数据对,则一共可以得到25个备选数据对,即可以确定出25个近邻数据对。最后,型号识别服务器01可以对该25个备选数据对按照与目标MAC地址的相似度由高到低的顺序进行排序,并确定出与目标MAC地址的相似度最高的10个备选数据对,该10个备选数据对中的10个MAC地址即为最终确定出的备选MAC地址,即与该目标MAC地址最近邻的10个MAC地址。

[0120] 应理解的是,相比于与传统的数据库查询方式,本申请实施例所采用的VP-tree算法可以大大降低数据组内查询的复杂度。譬如,查询数据组内与目标MAC地址最相似的2个数据对时,传统的多路搜索树(B-tree)算法的计算复杂度为 $O(2 * \log(n))$ ,其中n数据组中包括的数据对的个数。而采用VP-tree算法时,由于MAC地址相似的数据对在数据组中的位置也接近(即索引2接近),因此型号识别服务器01通过计算目标MAC地址与VP-tree节点的度量距离(即与中心数据对中的MAC地址的相似度)而迅速定位到最近邻点之后,就可以根据最近邻点在VP-tree中的位置的附近小范围内快速定位到其余的近邻点。因此采用VP-tree算法的计算复杂度可以降低至 $O(\log(n))$ 。

[0121] 示例的,如图7所示,假设型号识别服务器01通过计算目标MAC地址与索引1为003,索引2为001的中心数据对的MAC地址的相似度后,确定该索引1为003的数据组为备选数据组。该索引1为003的数据组的中心数据对的索引2为001,则型号识别服务器01可以在与该索引1为003,索引2为001的中心数据对的附近,快速定位到MAC地址与该目标MAC地址相似的其他数据对。例如图7中所示,由于构建VP-tree时,索引2为002的数据对,以及索引2为003的数据对与该中心数据对的相似度较高,因此该两个数据对在VP-tree中位于该中心数

据对的附近,进而型号识别服务器01可以快速确定出该索引1为003的备选数据组中,与目标MAC地址近邻的其他数据对为索引2为002的数据对,以及索引2为003的数据对。

[0122] 步骤106、将该第一数量个备选MAC地址对应的设备型号中,出现次数最多的设备型号确定为该目标终端设备的设备型号。

[0123] 型号识别服务器01确定出第一数量个备选MAC地址后,可以确定每个备选MAC地址对应的设备型号,并统计每个设备型号的出现次数,即重复次数。之后,型号识别服务器01可以将出现次数最多的设备型号确定为该目标终端设备的设备型号。

[0124] 需要说明的是,为了确保最终确定出的设备型号的可靠性,型号识别服务器01中可以预先存储有相似度阈值,型号识别服务器01在确定出第一数量个备选MAC地址后,还可以先剔除与该目标MAC地址的相似度小于该相似度阈值的备选MAC地址,然后再从剩余的备选MAC地址对应的设备型号中,确定该目标终端设备的设备型号。若每个备选MAC地址与该目标MAC地址的相似度均小于该相似度阈值,则型号识别服务器01可以确定该目标终端设备的设备型号为未知型号,即设备型号为“unknown”。

[0125] 还需要说明的是,在上述步骤101中,若型号识别服务器01获取到了多个目标终端设备的目标MAC地址和目标设备名称,则对于每个目标终端设备,型号识别服务器01均可以基于上述步骤102至106所示的方法识别该目标终端设备的设备型号。例如参考图2,型号识别服务器01可以根据设备名称N1和MAC地址M1确定出设备型号X1,并可以根据设备名称N2和MAC地址M2确定出设备型号X2。

[0126] 还需要说明的是,本申请实施例提供的设备型号的识别方法的步骤先后顺序可以进行适当调整,步骤也可以根据情况进行相应增减。例如,上述步骤101中可以仅获取目标终端设备的目标MAC地址,相应的,步骤102至步骤104可以根据情况删除。也即是,型号识别服务器01可以根据目标MAC地址与数据库中存储的MAC地址的相似度,识别目标终端设备的设备型号。由此,即使数据库中未存储该目标MAC地址对应的设备型号,也可以根据MAC地址的相似度确定出该目标终端设备的设备型号,从而有效提高了设备型号识别的成功率,降低了对数据库中存储的数据量的要求。任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化的方法,都应涵盖在本申请的保护范围之内,因此不再赘述。

[0127] 综上所述,本申请实施例提供了一种设备型号的识别方法,该方法可以结合设备名称和MAC地址两个参数来确定目标终端设备的设备型号,相比于仅基于MAC地址或者仅基于设备名称确定设备信号,可以有效提高设备型号识别的成功率和可靠性。

[0128] 在本申请实施例中,该型号识别服务器01中还可以存储有型号识别模型,该型号识别模型基于已确定设备型号的多个MAC地址样本训练得到。因此型号识别服务器01在基于目标MAC地址确定设备型号时,除了可以采用上述步骤105和步骤106所示的查询数据库的方法,还可以直接将该目标MAC地址输入至该型号识别模型,从而得到该型号识别模型输出的设备型号。

[0129] 可选的,型号识别服务器01可以先对目标终端设备的目标MAC地址进行one-hot编码,得到该目标MAC地址的地址向量。然后将该地址向量输入至该型号识别模型,该型号识别模型即可输出该目标终端设备属于不同设备型号的概率。型号识别服务器01进而可以将概率最高的设备型号确定为该目标终端设备的设备型号。若该型号识别模型输出的该目标终端设备属于每个设备型号的概率均小于概率阈值(例如0.8),则型号识别服务器01

可以确定该目标终端设备的设备型号为未知型号,例如可以将该目标终端设备的设备型号标注为unknown。

[0130] 示例的,如图8所示,假设型号识别服务器01获取到的目标终端设备的目标MAC地址为00CDFE8C1ACE,则型号识别服务器01对该目标MAC地址进行one-hot编码后,可以得到地址向量[0,0,12,13,15,14,8,12,1,10,12,14]。之后型号识别服务器01可以将该地址向量输入至型号识别模型。若该型号识别模型输出的目标终端设备属于设备型号EE-e1的概率最高,则型号识别服务器01可以确定该目标终端设备的设备型号为EE-e1。

[0131] 下文对该型号识别服务器01中预先存储的型号确定模型、相似度模型、数据库以及型号识别模型的构建过程进行介绍。

[0132] 图9是本申请实施例提供的一种型号确定模型的训练方法的流程图,该方法可以应用于型号识别服务器01,例如可以应用于图1所示系统中的第二服务器012。下文以该训练方法应用于第二服务器012为例进行说明。如图9所示,该方法可以包括:

[0133] 步骤201、获取设备名称样本以及该设备名称样本对应的设备型号样本。

[0134] 在本申请实施例中,第二服务器012可以获取大量标签数据,每个标签数据可以包括一个设备名称样本以及该设备名称样本对应的设备型号样本。也即是,第二服务器012可以获取到大量已经确定了设备型号的设备名称样本。

[0135] 步骤202、将该设备名称样本中,与该设备型号样本匹配的字符串中的每个字符均标注为有效字符,将除该字符串之外的其他字符均标注为无效字符。

[0136] 第二服务器012可以根据设备型号样本对设备名称样本进行字符级的标注,得到标识向量样本。该标识向量样本包括多个标识符,每个标识符可以用于指示该设备名称样本中的一个字符是否为有效字符。例如标识符的取值为第一整数时可以表示其所指示的字符为有效字符,标识符的取值为第二整数时可以表示其所指示的字符为无效字符。通过对设备名称样本中的每个字符进行标注,也即是实现了对设备名称样本中的字符的分类。

[0137] 示例的,参考图5,假设第二服务器012获取到的设备名称样本为“小明的AA-a1-666”,对应的设备型号样本为“AA-a1”,并且该第一整数包括0和1(其中0表示第一个有效字符,1表示位于中间的有效字符),第二整数为2。则第二服务器012可以确定设备名称样本中与该设备型号样本匹配的字符串为“AA-a1”,并可以将该字符串中的每个字符均标注为有效字符,即确定该每个字符的标识符均为第一整数。相应的,对于该设备名称样本中除字符串“AA-a1”之外的其他字符,则均可以标注为无效字符,即确定其他每个字符的标识符均为第二整数。如图5所示,第二服务器012对设备名称样本“小明的AA-a1-666”进行标注后得到的标识向量样本可以为[2,2,2,0,1,1,1,1,1,2,2,2],

[0138] 步骤203、对标注后的设备名称样本进行模型训练,得到型号确定模型。

[0139] 第二服务器012可以采用NLP技术对标注后的设备名称样本进行模型训练,得到型号确定模型。

[0140] 可选的,第二服务器012还需对设备名称样本进行编码,得到机器可识别的名称向量样本。例如,第二服务器012可以采用one-hot编码算法对该设备名称样本进行编码,从而将该设备名称样本中的每个字符均转换为机器可识别的数值。之后,第二服务器012即可将该名称向量样本作为模型的输入,并将该标识向量样本作为模型的输出进行模型训练,得到该型号确定模型。

[0141] 在本申请实施例中,该第二服务器012可以采用双向长短期记忆(Bi-directional long short term memory,Bi-LSTM)算法和条件随机场(conditional random fields,CRF)算法构建型号确定模型。该Bi-LSTM算法可以提取每个字符的上下文语义,CRF算法再对Bi-LSTM算法的处理结果选出合理的分类结果。通过上述方式,可以训练出较为优化的具有识别设备型号功能的神经网络模型,即型号确定模型。

[0142] 其中,该模型训练过程还可以采用交叉验证机制,即第二服务器012可以将多个标签数据分成训练集和验证集。在神经网络模型的每一轮训练中,第二服务器012采用训练集中的标签数据集更新模型参数,再采用验证集中的标签数据验证更新结果。如此反复训练,直至模型最优。

[0143] 图10是本申请实施例提供的一种相似度模型的训练方法的流程图,该方法可以应用于型号识别服务器01,例如可以应用于图1所示系统中的第二服务器012。下文以该训练方法应用于第二服务器012为例进行说明。如图10所示,该方法可以包括:

[0144] 步骤301、对获取到的多个数据对进行预处理。

[0145] 在本申请实时中,第二服务器012可以获取到网管设备03下发的用于训练相似度模型的多个数据对,每个数据对均包括一个MAC地址,以及与该MAC地址对应的设备型号。其中,每个数据对中的设备型号可以是采用型号确定模型对设备名称进行处理得到的。

[0146] 可选的,为了确保获取到的数据对的可靠性,第二服务器012可以先对获取到的数据对进行预处理,以清洗掉一些脏数据。其中脏数据是指设备型号与MAC地址给出的厂商信息明显不吻合的数据。由于MAC地址通常由12个16进制整数(0-9,a-f)组成,其前六位为组织唯一标识符(organization unique identifier,OUI),该OUI通常用于标识设备的网卡生产商,因此第二服务器012可以通过MAC地址中的OUI识别厂商信息。

[0147] 其中,设备型号与厂商信息不吻合可能有两个原因:(1)设备名称有误:用户可能会随意修改设备名称,比如将AA品牌的手机的设备名称修改为BB品牌;(2)MAC地址有误:某些终端设备会在连接网关时修改MAC地址,从而防止跟踪。

[0148] 在本申请实施例中,在清洗脏数据时,第二服务器012可以检测根据设备型号确定出的设备厂商,与根据MAC地址确定出的设备厂商是否一致。若一致则保留数据对,若不一致则可以确定该数据对为脏数据,因此可以删除。

[0149] 步骤302、根据预处理后的数据对,确定多个训练样本。

[0150] 第二服务器012在完成数据对的清洗后,可以从预处理后的数据对包括的MAC地址中确定出多个MAC地址组,每个MAC地址组可以包括至少两个MAC地址。并且,对于每个MAC地址组,第二服务器012可以获取该MAC地址组中任意两个MAC地址的相似度,该相似度可以是人工标注的,也可以是第二服务器012根据该两个MAC地址所对应的设备型号的相似程度自动标注的。

[0151] 第二服务器012获取到每个MAC地址组中任意两个MAC地址的相似度后,即可基于该多个MAC地址组以及获取到的相似度生成多个训练样本。其中,每个训练样本包括至少两个MAC地址样本,以及该至少两个MAC地址样本中任意两个MAC地址样本的相似度。

[0152] 可选的,对于每个MAC地址样本,第二服务器012可以对其进行编码,得到机器可识别的地址向量样本。例如,第二服务器012可以对每个MAC地址样本中的每个字符进行one-hot编码。或者,也可以采用其他编码方式,比如也可以对MAC地址的前六位和后六位分别采

用不同的方式编码。

[0153] 对于任意两个MAC地址样本的相似度,标注人员或第二服务器012可以根据该两个MAC地址样本对应的设备型号确定该两个MAC地址样本的相似度 $y$ 的取值,且 $y$ 的取值可以与两个MAC地址样本的相似度负相关。例如,当两个MAC地址样本所属的终端设备为不同厂商的设备时,相似度 $y$ 的取值可以为: $y=2$ ,表示该两个MAC地址样本的相似度很低。当两个MAC地址样本所属的终端设备属于同一厂商但设备型号不同时,相似度 $y$ 的取值可以为: $y=1$ ,表示该两个地址样本的相似度较高。当两个MAC地址样本所属的终端设备属于同一厂商,且设备型号相同时,相似度 $y$ 的取值可以为: $y=0$ ,表示该两个地址样本的相似度很高。

[0154] 示例的,假设两个MAC地址样本所属的终端设备分别为:AA品牌的型号为a1的手机(即设备型号为AA-a1),以及AA品牌的型号为a2的手机(即设备型号为AA-a2),则第二服务器012可以确定该两个MAC地址样本的相似度为: $y=1$ 。

[0155] 步骤303、对该多个训练样本进行训练,得到相似度模型。

[0156] 第二服务器012可以将每个训练样本中MAC地址样本的地址样本向量作为输入,将任意两个地址样本的相似度作为目标(target)输出,并采用深度度量学习(deep metric learning)算法进行训练模型,得到相似度模型。

[0157] 基于该深度度量学习算法可以学习到的相似度模型可以理解为用于衡量MAC地址之间的相似度(也可以称为距离或度量距离)的函数,该相似度模型可以使得相同设备型号的终端设备的MAC地址的相似度较高(即度量距离较小),而不同设备型号的终端设备的MAC地址的相似度较低(即度量距离较大)。

[0158] 示例的,假设网管设备03提供了1000个数据对,第二服务器012可以基于该1000个数据对生成 $1000 \times 500 = 500000$ 个不同的MAC地址组,每个MAC地址组包括两个MAC地址。之后,第二服务器012可以随机选取100000个MAC地址组,以及每个MAC地址组中两个MAC地址的相似度,得到100000个训练样本。第二服务器012对该100000个训练样本进行训练,即可得到相似度模型。对于任意输入的两个MAC地址,该相似度模型均可以输出一个相似度 $y$ ,该相似度 $y$ 可以衡量输入的两个MAC地址对应的设备型号是否相似。

[0159] 图11是本申请实施例提供的一种数据库的构建方法的流程图,该方法可以应用于型号识别服务器01,例如可以应用于图1所示系统中的第二服务器012。下文以该构建方法应用于第二服务器012为例进行说明。如图11所示,该方法可以包括:

[0160] 步骤401、获取多个数据对。

[0161] 在本申请实时中,第二服务器012可以获取到网管设备03下发的用于构建数据库的多个数据对,每个数据对均包括一个MAC地址,以及与该MAC地址对应的设备型号。其中,用于构建数据库的多个数据对与上述步骤301中用于训练相似度模型的多个数据对可以相同,也可以不同,本申请实施例对此不做限定。

[0162] 为了确保获取到的数据对的可靠性,第二服务器012可以先对获取到的数据对进行预处理,以清洗掉一些脏数据。清洗脏数据的过程可以参考上述步骤301,此处不再赘述。

[0163] 步骤402、采用聚类算法对该多个数据对进行分组,得到多个数据组。

[0164] 第二服务器012可以根据相似度模型,构建基于LSH技术的数据库,使得海量高维的数据(MAC地址经历了one-hot编码以及深度度量学习会映射成高维的矢量)的快速近邻搜索成为了可能。

[0165] 大数据表明,设备型号对应的MAC地址大致具有分段连续的特征,即厂家习惯把一段连续的MAC地址分配给同一设备型号的终端设备,即同一设备型号的终端设备的MAC地址通常占据一个MAC地址区间。然而由于终端设备的型号太多,这种分段连续规律又极其复杂,例如某一型号的手机可能会占据多个MAC地址区间。随着采集到的数据的增多,一方面这种区间连续的规律越来越难以用规则去描述,另一方面基于传统索引方式的数据库会使得查询效率变得很低。因此本申请实施例提供了基于相似度模型和LSH技术的数据库创建方案。

[0166] 可选的,第二服务器012可以先通过相似度模型确定不同MAC地址之间的相似度,进而可以通过聚类算法将相似度较高的MAC地址聚类在一起,即划分至同一个数据组。由此,即使某个型号的手机的MAC地址占据了多个MAC地址区间,经过基于相似度的聚类之后,该型号的手机的所有MAC地址也可以被划分至同一个数据组。

[0167] 在本申请实施例中,第二服务器012可以采用k-centers这一聚类算法(当然也可以采用其他基于距离的聚类算法),将获取到的数据对划分成N个数据组,即N个cluster,每个cluster可以包括一个或多个数据对。并且,不同数据组包括的数据对的个数可以相同,也可以不同。其中,N可以为预先设定的大于1的整数,例如N可以等于该多个数据对中包括的设备型号的数量,或者等于设备型号的数量的10倍。

[0168] 采用k-centers算法进行聚类的过程如下:

[0169] 步骤S11、从多个数据对中任意选取一个数据对作为第一个数据组的聚类中心,即中心数据对。

[0170] 步骤S12、计算剩余的每个数据对中的MAC地址与该第一个数据组的中心数据对中的MAC地址的相似度,将相似度最小(即度量距离最大)的数据对作为第二个数据组的中心数据对。

[0171] 步骤S13、继续计算剩余的每个数据对的MAC地址与该第一个数据组的中心数据对的MAC地址的相似度,以及与该第二个数据组的中心数据对的MAC地址的相似度,并将相似度最小(即度量距离最大)的数据对作为第三个数据组的中心数据对。

[0172] 以此类推,直至确定出N个数据组的中心数据对。其中,该步骤S13中所述的“相似度最小”可以是指:与该两个中心数据对的MAC地址的相似度之和最小,或者相似度的均值最小。

[0173] 步骤S14、对于除中心数据对之外的每个数据对,确定该数据对的MAC地址与每个中心数据对的MAC地址的相似度,并将该数据对划分至相似度最高的中心数据对所属的数据组。

[0174] 例如,假设某数据对的MAC地址与第一个数据组的中心数据对的MAC地址的相似度最高,则可以将该数据对划分至第一个数据组。

[0175] 通过上述方法,第二服务器012即可将获取到的数据对划分成N个数据组。并且,通过k-centers算法可以确保划分至同一个数据组的各个数据对的MAC地址较为相似,即MAC地址之间的度量距离较小。而属于不同数据组的两个数据对的MAC地址的相似度则较低,即两个数据对的MAC地址的度量距离较大。由此,对于待查询的目标MAC地址,可以通过匹配该目标MAC地址与各个中心数据对的MAC地址的相似度,快速锁定目标MAC地址的近邻(即与目标MAC地址相似度较高的MAC地址)所在的区域。

[0176] 示例的,假设第二服务器012获取到了如图12所示的多个数据对,则该第二服务器012采用相似度模型以及聚类算法对该多个数据对进行聚类后,可以得到c1至c7共7个数据组,且每个数据组中各个数据对的设备型号可以相同。

[0177] 需要说明的是,每个数据对在数据库中的存储位置可以用两级索引(索引1,索引2)表示。其中,每个数据对的索引1可以为其所属的数据组的索引,即同一数据组中的各个数据对的索引1相同。例如,结合图7和图12,数据组c1中每个数据对的索引1可以均为001,数据组c2中每个数据对的索引1可以均为002。

[0178] 步骤403、对于每个数据组,根据数据组中的中心数据对的MAC地址与其他每个数据对中的MAC地址的相似度,对该数据组包括的数据对按照相似度由高到低的顺序进行排序。

[0179] 为了有效提高每个数据组内的MAC地址的近邻查询效率,第二服务器012还可以对每个数据组中的数据对按照与该中心数据对的MAC地址的相似度由高到低的顺序进行排序。之后,第二服务器012即可为该排序后的各个数据对分配索引2。

[0180] 可选的,第二服务器012可以采用VP-tree算法对每个数据组中的各个数据对进行排序。其中,对任一数据组中的数据对的排序过程如下:

[0181] 步骤S21、将该数据组中的中心数据对确定为VP-tree的根(root)节点。

[0182] 步骤S22、计算数据组内其他每个数据对的MAC地址与该根节点的MAC地址的相似度,根据计算得到的相似度的中位数将除根节点之外的其他数据对划分为两个子集。

[0183] 该两个子集中的一个子集可以包括:与该根节点的MAC地址的相似度大于或等于该中位数的数据对,该子集即为VP-tree的左子树。另一个子集可以包括:与该根节点的MAC地址的相似度小于该中位数的数据对,该子集即为VP-tree的右子树。

[0184] 步骤S23、对于每个子集,从该子集中选取一个数据对作为该子集的新的子节点,计算该子集中的其他数据对的MAC地址与该子节点的MAC地址的相似度,进而根据计算得到的相似度的中位数将除子节点之外的其他数据再次拆分成两个子集。其中,该新的子节点可以为该子集中的任意一个数据对;或者也可以是该子集中各个数据对与该根节点的MAC地址的相似度的中位数所对应的一个数据对;又或者,还可以是该子集中与该根节点的MAC地址的相似度最高的一个数据对。

[0185] 以此类推,直至每个子集均只剩下一个数据对,由此即可完成VP-tree的构建,即数据对的排序。该VP-tree中各个数据对可以按照从上至下(即从根节点至尾节点)以及从左至右(即从左子树至右子树)的顺序排列。

[0186] 示例的,参考图13,假设某个数据组中包括D1至D7共7个数据对,其中数据对D1为中心数据对,则第二服务器012以该中心数据对D1为根节点构造的VP-tree可以如图13所示。该VP-tree中的7个数据对按照D1、D3、D2、D6、D5、D4和D7的顺序排列。

[0187] 根据上文描述可知,基于LSH技术构建的数据库,可以使得相似设备型号的数据对在数据库中的位置也相近,从而便于近邻搜索。

[0188] 本申请实施例还提供了一种型号识别模型的训练方法,该方法可以包括如下步骤:

[0189] 步骤S31、第二服务器012获取多个MAC地址样本,以及每个MAC地址样本对应的设备型号。

[0190] 该多个MAC地址样本,以及每个MAC地址样本对应的设备型号可以是网管设备03下发至第二服务器012的。

[0191] 步骤S32、第二服务器对该多个MAC地址样本,以及每个MAC地址样本对应的设备型号进行训练,得到型号识别模型。

[0192] 在本申请实施例中,第二服务器012可以先对获取到的数据进行预处理,例如去除脏数据。然后对预处理后得到的每个MAC地址样本进行编码。之后,第二服务器012可以将该编码后的MAC地址样本作为模型的输入,将MAC地址样本对应的设备型号作为模型的目标输出,采用深度学习或随机森林等算法进行模型训练,直至损失函数(loss function)收敛即可得到型号识别模型。由于该型号识别模型可以输出终端设备属于多个设备型号中每个设备型号的概率,因此该型号识别模型也可以称为分类器(classifier)。

[0193] 在本申请实施例中,型号识别服务器01在识别出目标终端设备的设备型号后,还可以将识别出的设备型号发送至网管设备03。网络管理员可以对该设备型号进行人工验证,若该设备型号错误,则网络管理员可以向网管设备03输入纠正的设备型号。并且,该网管设备03还可以定期将收集到的纠正数据发送至型号识别服务器01,以便该型号识别服务器01可以基于该纠正数据对模型(例如相似度模型、型号确定模型以及型号识别模型)进行重训练。其中,该纠正数据可以包括设备名称、MAC地址以及纠正的设备型号;或者可以仅包括MAC地址和纠正的设备型号。

[0194] 示例的,网管设备03可以将纠正数据发送至第二服务器012,以触发该第二服务器012进行模型的重训练。该第二服务器012完成重训练后,可以将更新的模型发送至第一服务器011,以便该第一服务器011可以基于更新后的模型进行设备型号的识别。通过该重训练机制,可以确保模型的不断完善和优化,提高设备型号识别的准确率。

[0195] 本申请实施例提供的方法,可以结合MAC地址和设备名称两个参数确定设备型号。对于能够提供有效的设备型号的设备名称,可以将基于该设备名称确定出的设备型号作为终端设备的设备型号,而对于无法提供有效设备型号的设备名称,则可以基于MAC地址来确定设备型号,由此有效提升了设备型号的识别率。例如,可以将识别率由55%提升至95%。其中,识别率是指识别出设备型号的终端设备的数量占全体通过该型号识别服务器识别设备型号的终端设备的比例。

[0196] 其中,在根据设备名称识别设备型号时,采用了基于NLP的型号确定模型,该模型能够自动根据标签数据学习设备型号的提取规律,相对于传统的人工标注或者基于复杂的规则的正则表达式的标注方法,开发代价小,模型通用性高,便于泛化到不同地域或者不同语系,模型的开发和维护成本较低。

[0197] 在根据MAC地址识别设备型号时,采用了深度度量学习以及LSH技术,实现了对设备型号在MAC地址上的分布规律的统计,从而从大数据和统计学的角度给出了关于设备型号的预测,使得高维的海量数据的快速近邻搜索成为了可能。

[0198] 综上可知,本申请实施例提供的方案主要具有如下优势:(1)基于MAC地址确定设备型号的方法以及基于设备名称确定设备型号的方法均是基于大数据和统计学的原理,识别准确率可以得到保证。(2)两种方法都是数据驱动的,不依赖于规则,方法容易泛化到不同的地域和不同的人文环境;模型维护代价小,数据增多之后模型只需要简单地重训练即可。(3)两个方法所采用的模型都很小,识别速度快,识别效率较高。(4)两个方法的算法理

论框架一致,可以复用,从而能够有效降低开发工作量。(5)具有重训练机制,能够有效增强模型预测的准确性和鲁棒性。

[0199] 图14是本申请实施例提供的一种设备型号的识别装置的结构示意图,如图14所示,该装置可以包括:

[0200] 第一获取模块501,用于获取目标终端设备的目标物理地址。该第一获取模块501的功能实现可以参考上述步骤101的相关描述。

[0201] 第一确定模块502,用于从数据库中确定第一数量个备选物理地址,该数据库包括多个物理地址,以及每个物理地址对应的设备型号,其中每个备选物理地址与该目标物理地址的相似度,大于该数据库中其他物理地址与该目标物理地址的相似度,该第一数量为大于1的整数。该第一确定模块502的功能实现可以参考上述步骤105的相关描述。

[0202] 第二确定模块503,将该第一数量个备选物理地址对应的设备型号中,出现次数最多的设备型号确定为该目标终端设备的设备型号。该第二确定模块503的功能实现可以参考上述步骤106的相关描述。

[0203] 可选的,该数据库包括多个数据组,每个数据组包括一个或多个数据对,每个数据对包括一个物理地址,以及与该物理地址对应的设备型号;该第一确定模块502可以用于:

[0204] 根据该目标物理地址与每个数据组中任一数据对的物理地址的相似度,确定第二数量个备选数据组,每个备选数据组中该任一数据对的物理地址与该目标物理地址的相似度,大于其他数据组中该任一数据对的物理地址与该目标物理地址的相似度,其中,该第二数量为大于1的整数;

[0205] 从该第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址,每个备选物理地址与该目标物理地址的相似度,大于该第二数量个备选数据组中的其他物理地址与该目标物理地址的相似度。

[0206] 可选的,每个数据组中存在一个中心数据对,且每个数据组中的数据对按照与该中心数据对的物理地址的相似度由高到低的顺序排列;每个数据组中的任一数据对为该中心数据对;

[0207] 该第一确定模块502可以用于:根据每个备选数据组中的数据对的排列顺序,从该第二数量个备选数据组包括的物理地址中确定出第一数量个备选物理地址。

[0208] 该第一确定模块502的功能实现还可以参考上述步骤1051和步骤1052的相关描述。

[0209] 可选的,如图15所示,该装置还可以包括:

[0210] 第二获取模块504,用于获取该多个数据对。该第二获取模块504的功能实现可以参考上述步骤401的相关描述。

[0211] 聚类模块505,用于采用聚类算法对该多个数据对进行分组,得到该多个数据组。该聚类模块505的功能实现可以参考上述步骤402的相关描述。

[0212] 排序模块506,用于对于每个数据组,根据该数据组中的中心数据对的物理地址与其他每个数据对中的物理地址的相似度,对该数据组包括的数据对按照相似度由高到低的顺序进行排序。该排序模块506的功能实现可以参考上述步骤403的相关描述。

[0213] 可选的,如图15和图16所示,该装置还可以包括:

[0214] 第三确定模块507,用于在该从数据库中确定第一数量个备选物理地址之前,采用

相似度模型确定该数据库中的物理地址与该目标物理地址的相似度;其中,该相似度模型基于已确定相似度的多个物理地址样本训练得到。

[0215] 可选的,该第一获取模块501,还可以用于获取该目标终端设备的目标设备名称。继续参考图15和图16,该装置还可以包括:

[0216] 第四确定模块508,用于根据该目标设备名称确定该目标终端设备的备选设备型号。该第四确定模块508的功能实现可以参考上述步骤102的相关描述。

[0217] 第五确定模块509,用于若该备选设备型号不为未知型号,则将该备选设备型号确定为该目标终端设备的设备型号。该第五确定模块509的功能实现可以参考上述步骤104的相关描述。

[0218] 相应的,该第一确定模块502,可以用于若该备选设备型号为该未知型号,则从数据库中确定第一数量个备选物理地址。

[0219] 可选的,该第四确定模块508可以用于:

[0220] 采用型号确定模型从该目标设备名称中确定该目标终端设备的备选设备型号;其中,该型号确定模型基于已确定设备型号的多个设备名称样本训练得到。

[0221] 可选的,该第四确定模块508可以用于:

[0222] 采用型号确定模型确定该目标设备名称中的每个字符是否为有效字符;将该目标设备名称中的有效字符组成的字符串确定为该目标终端设备的备选设备型号。该第四确定模块508的功能实现还可以参考上述步骤1021至步骤1023的相关描述。

[0223] 可选的,如图15所示,该装置还可以包括:

[0224] 第三获取模块510,用于获取设备名称样本以及该设备名称样本对应的设备型号样本。该第三获取模块510的功能实现可以参考上述步骤201的相关描述。

[0225] 第六确定模块511,用于将该设备名称样本中,与该设备型号样本匹配的字符串中的每个字符均标注为有效字符,将除该字符串之外的其他字符均标注为无效字符。该第六确定模块511的功能实现可以参考上述步骤202的相关描述。

[0226] 训练模块512,用于对标注后的该设备名称样本进行模型训练,得到该型号确定模型。该训练模块512的功能实现可以参考上述步骤203的相关描述。

[0227] 可选的,该第四确定模块508可以用于:

[0228] 分别确定该目标设备名称与多个设备型号模板中每个设备型号模板的匹配度;将匹配度最高的设备型号模板确定为该目标终端设备的备选设备型号。

[0229] 综上所述,本申请实施例提供了一种设备型号的识别装置,该装置可以根据目标MAC地址与数据库中存储的MAC地址的相似度,从数据库中确定出相似度较高的第一数量个备选MAC地址,然后再将该第一数量个备选MAC地址对应的设备型号中,出现次数最多的设备型号确定为目标终端设备的设备型号。由此,即使数据库中未存储该目标MAC地址对应的设备型号,也可以根据MAC地址的相似度确定出该目标终端设备的设备型号,从而有效提高了设备型号识别的成功率,降低了对数据库中存储的数据量的要求。

[0230] 并且,本申请实施例提供的装置,还可以结合设备名称和MAC地址两个参数来确定目标终端设备的设备型号,相比于仅基于MAC地址或者仅基于设备名称确定设备信号,可以有效提高设备型号识别的成功率和可靠性。

[0231] 应理解的是,本申请实施例提供的设备型号的识别装置可以用专用集成电路

(application-specific integrated circuit,ASIC)实现,或可编程逻辑器件(programmable logic device,PLD)实现,上述PLD可以是复杂程序逻辑器件(complex programmable logical device,CPLD),现场可编程门阵列(field-programmable gate array,FPGA),通用阵列逻辑(generic array logic,GAL)或其任意组合。

[0232] 图17是本申请实施例提供的一种设备型号的识别装置的结构示意图,参考图17,该设备型号的识别装置可以包括:处理器1701、存储器1702、网络接口1703和总线1704。其中,总线1704用于连接处理器1701、存储器1702和网络接口1703。通过网络接口1703(可以是有线或者无线)可以实现与其他器件之间的通信连接。存储器1702中存储有计算机程序17021,该计算机程序17021用于实现各种应用功能。

[0233] 应理解,在本申请实施例中,处理器1701可以是CPU,该处理器1701还可以是其他通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、GPU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是任何常规的处理器等。

[0234] 存储器1702可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(read-only memory,ROM)、可编程只读存储器(programmable ROM,PROM)、可擦除可编程只读存储器(erasable PROM,EPROM)、电可擦除可编程只读存储器(electrically EPROM,EEPROM)或闪存。易失性存储器可以是随机存取存储器(random access memory,RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器(static RAM,SRAM)、动态随机存取存储器(DRAM)、同步动态随机存取存储器(synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(double data rate SDRAM,DDR SDRAM)、增强型同步动态随机存取存储器(enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(synchlink DRAM,SLDRAM)和直接内存总线随机存取存储器(direct rambus RAM,DR RAM)。

[0235] 总线1704除包括数据总线之外,还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见,在图中将各种总线都标为总线1704。

[0236] 处理器1701被配置为执行存储器1702中存储的计算机程序,处理器1701通过执行该计算机程序17021来实现上述方法实施例中的步骤。

[0237] 本申请实施例还提供了一种计算机可读存储介质,该计算机可读存储介质中存储有指令,当该计算机可读存储介质在计算机上运行时,使得计算机执行如上述方法实施例中的步骤。

[0238] 本申请实施例还提供了一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述方法实施例中的步骤。

[0239] 本申请实施例还提供了一种设备型号的识别系统,如图1所示,该系统可以包括:第一服务器011和第二服务器012。

[0240] 该第一服务器011可以用于实现如图3、图4和图6所示的方法实施例中的步骤;该第二服务器012可以用于实现如图9至图11所示方法实施例中的步骤。

[0241] 示例的,该第一服务器011可以包括如图14或图16所示的装置。该第二服务器012可以包括如图15所示的装置中的模块504至506,以及模块510至512。

[0242] 如图1所示,该系统还可以包括:网关设备02。该网关设备02可以分别与终端设备

04和该第一服务器011连接,该网关设备02可以用于获取该终端设备04的MAC地址,并将获取到的MAC地址发送至该第一服务器011。

[0243] 可选的,本申请实施例提供的设备型号的识别系统也可以仅包括第一服务器011,该第一服务器011可以用于实现如图3、图4、图6以及图9至图11所示的方法实施例中的步骤。例如,该第一服务器011可以包括如图15所示的装置。

[0244] 应当理解的是,在本文中提及的“和/或”,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0245] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的模块及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0246] 本领域普通技术人员可以清楚地了解到,为了描述的方便和简洁,上述描述的系统、装置和模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0247] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个模块或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或模块的间接耦合或通信连接,也可以是电的,机械的或其它的形式连接。

[0248] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,即可以位于一个地方,或者也可以分布到多个网络模块上。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申请实施例方案的目的。

[0249] 另外,在本申请各个实施例中的各功能模块可以集成在一个处理模块中,也可以是各个模块单独物理存在,也可以是两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。

[0250] 所述集成的模块如果以软件功能模块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,本申请的技术方案本质上或者说对现有技术做出贡献的部分,或者该技术方案的全部或部分可以以计算机程序产品的形式体现出来,该计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时,全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。该计算机程序产品存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机

能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质, (例如, 软盘、硬盘、磁带)、光介质 (例如, DVD)、或者半导体介质 (例如固态硬盘Solid State Disk (SSD)) 等。

[0251] 以上所述, 仅为本申请的具体实施方式, 但本申请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申请揭露的技术范围内, 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申请的保护范围之内。因此, 本申请的保护范围应以权利要求的保护范围为准。

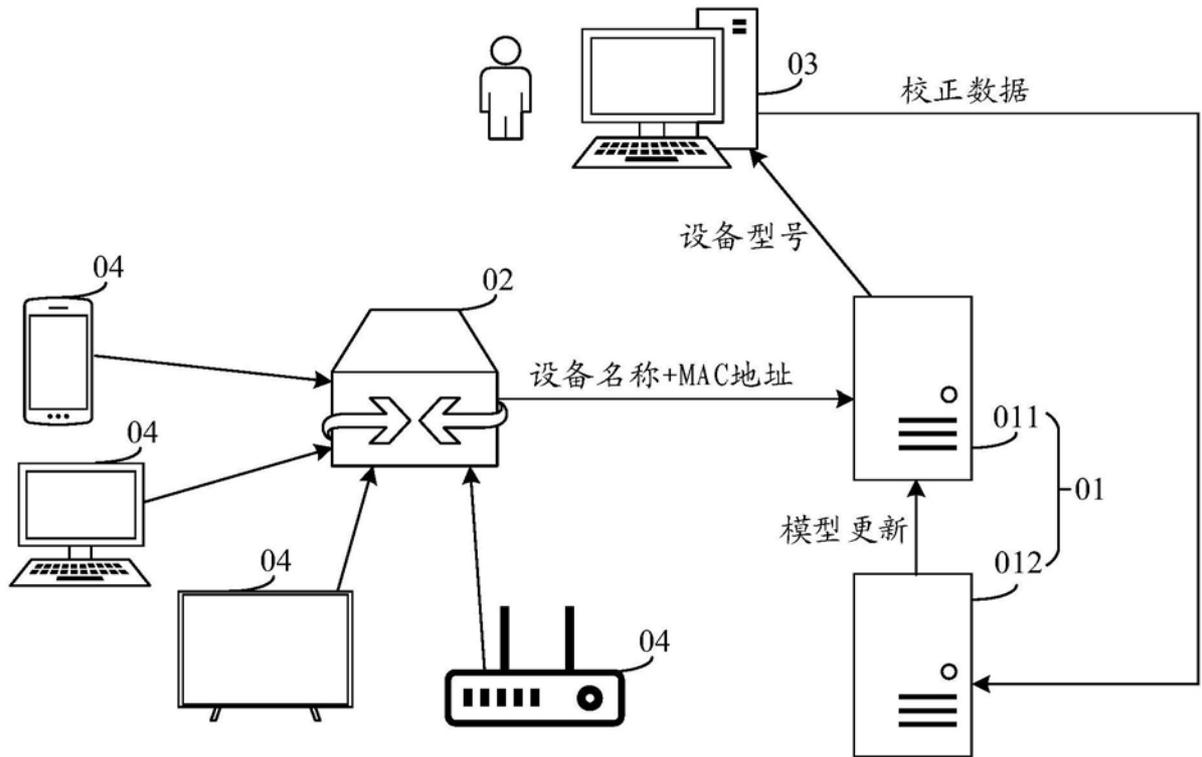


图1

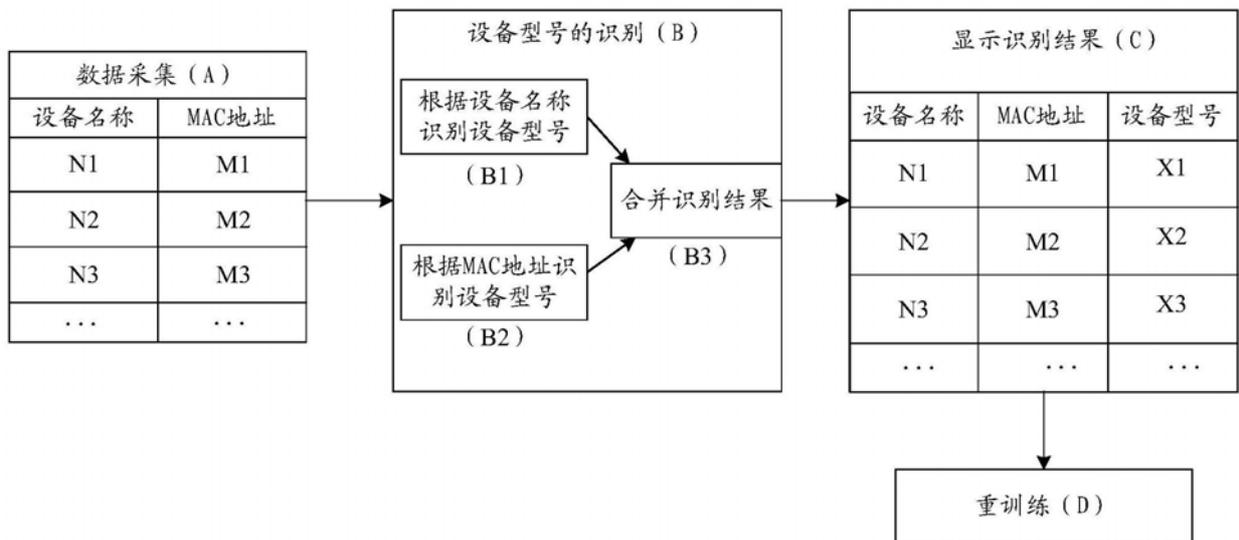


图2

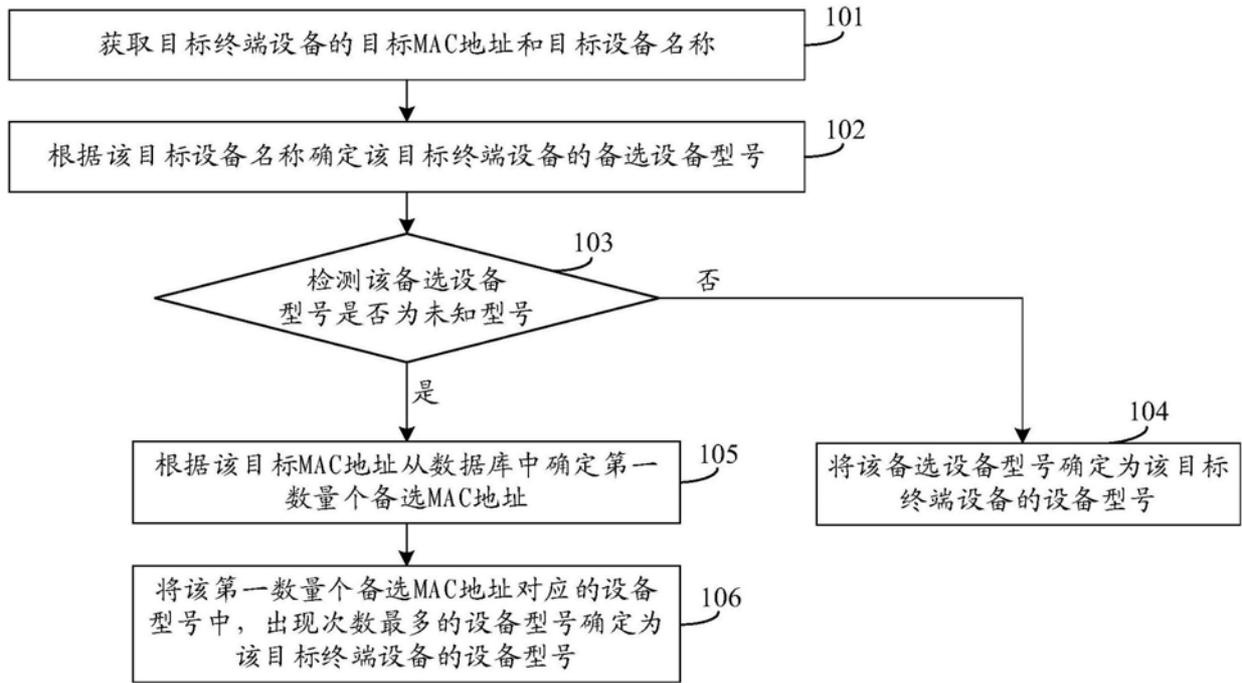


图3

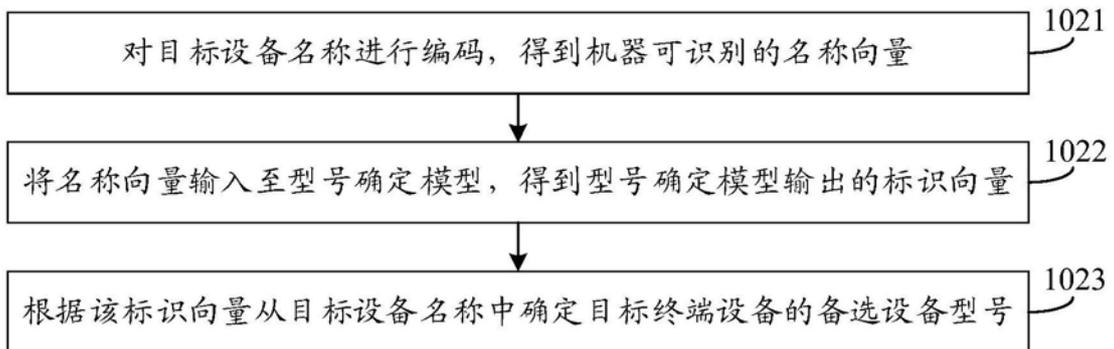


图4

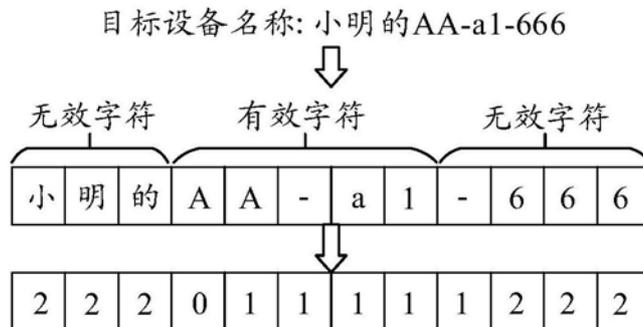


图5

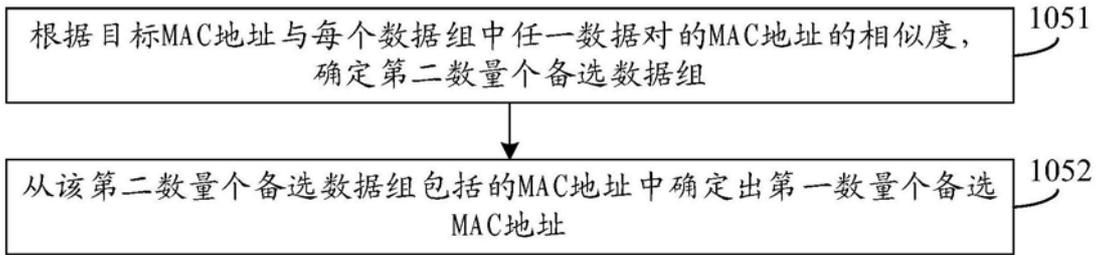


图6

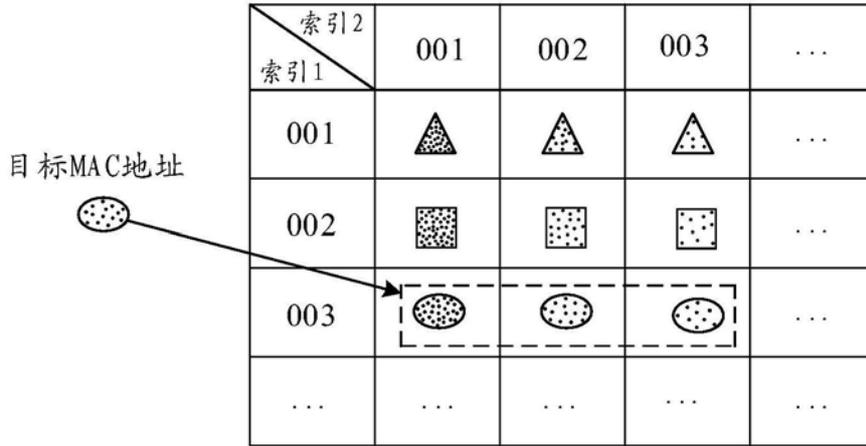


图7

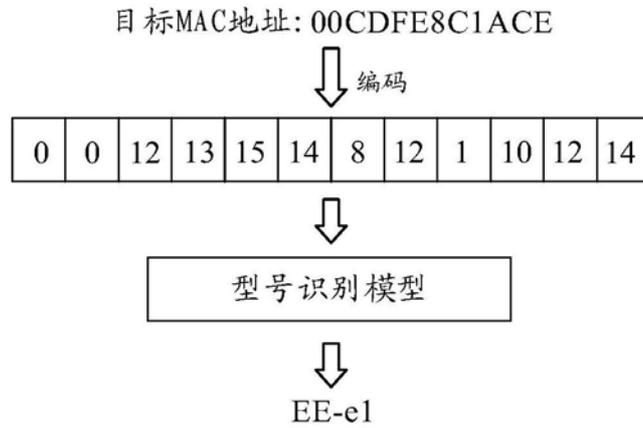


图8

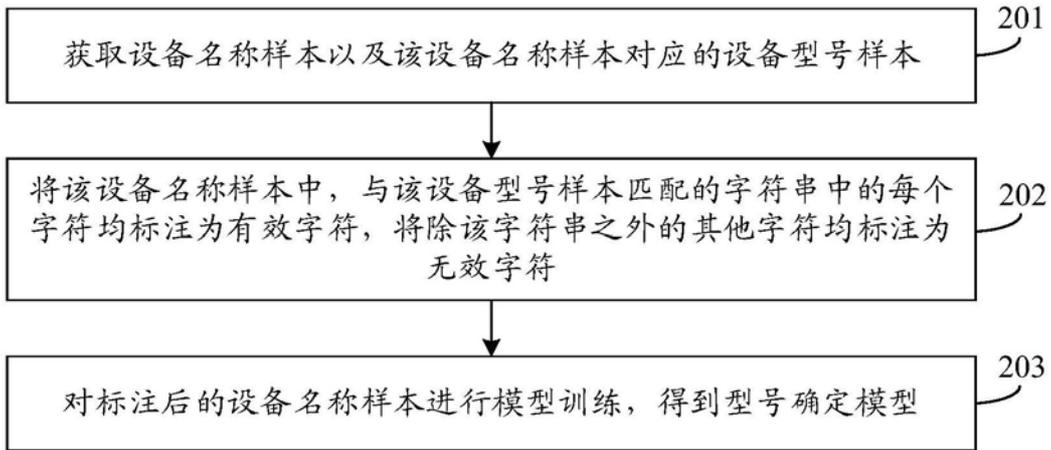


图9

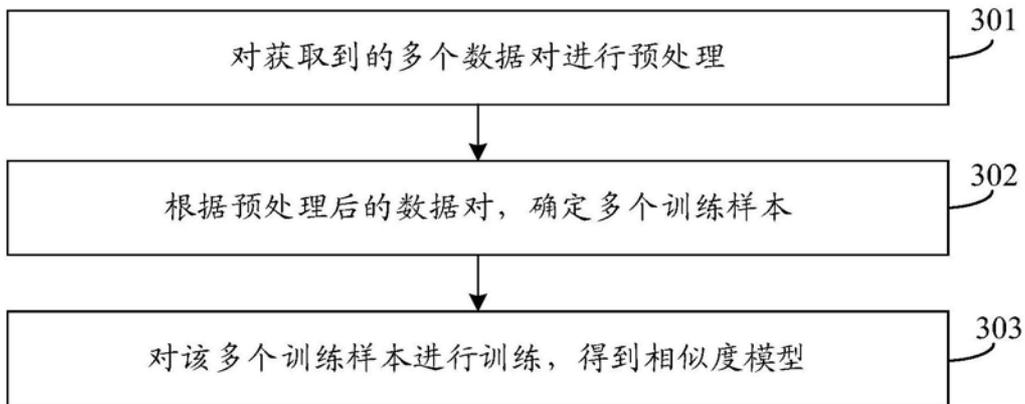


图10

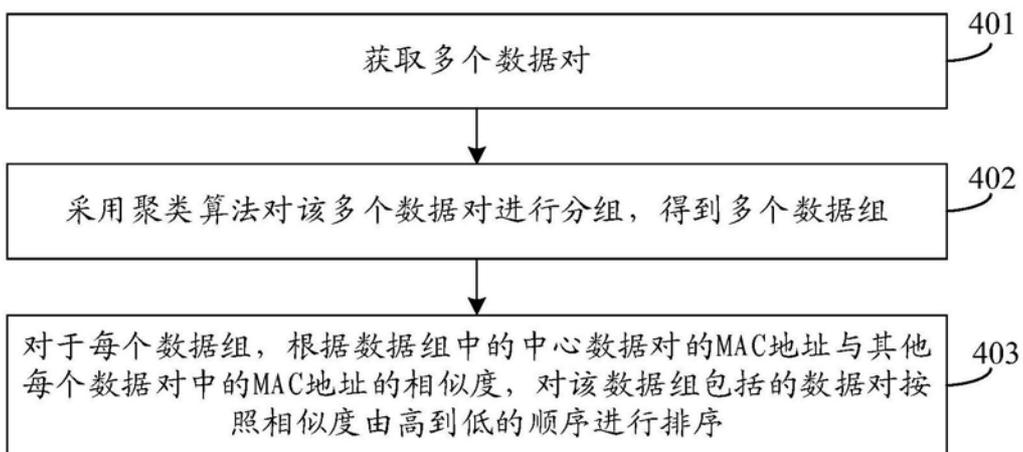


图11

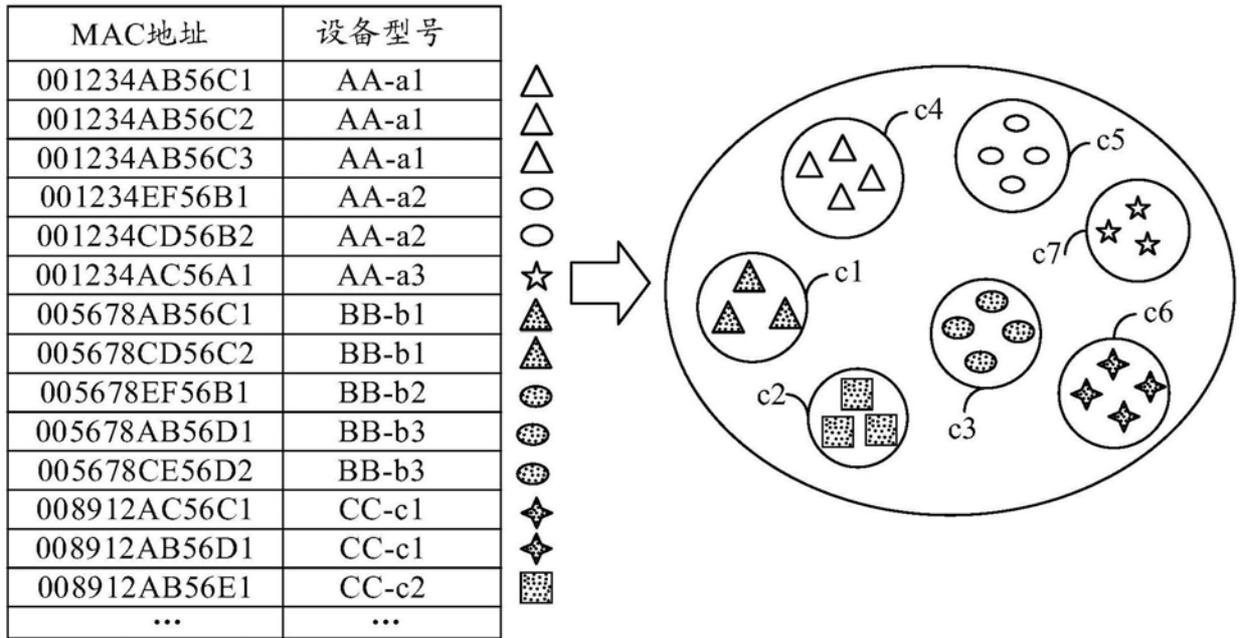


图12

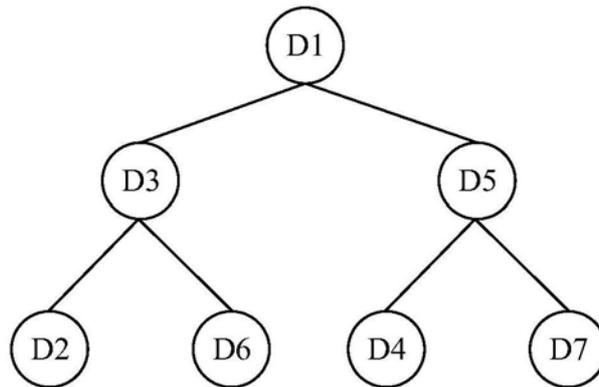


图13

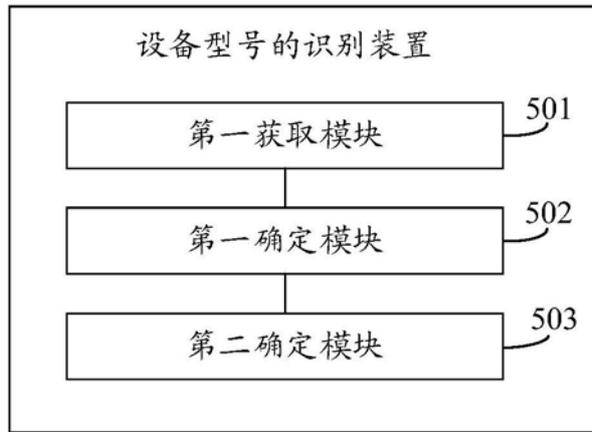


图14

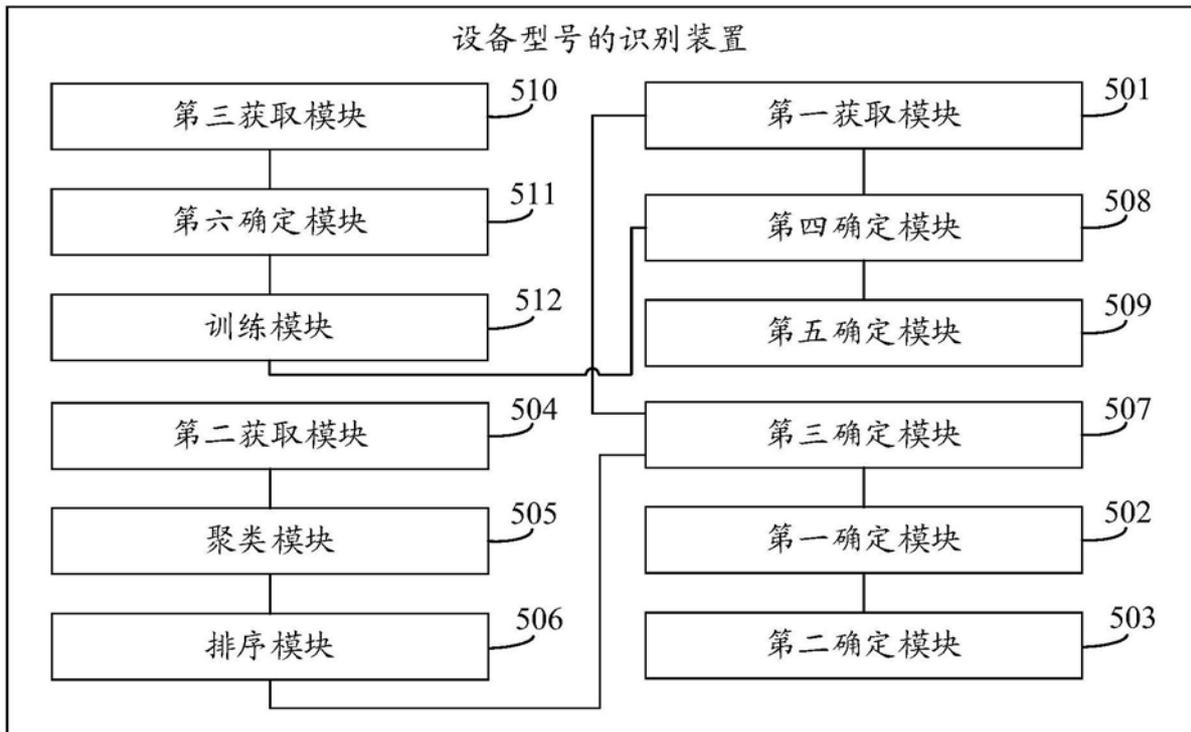


图15

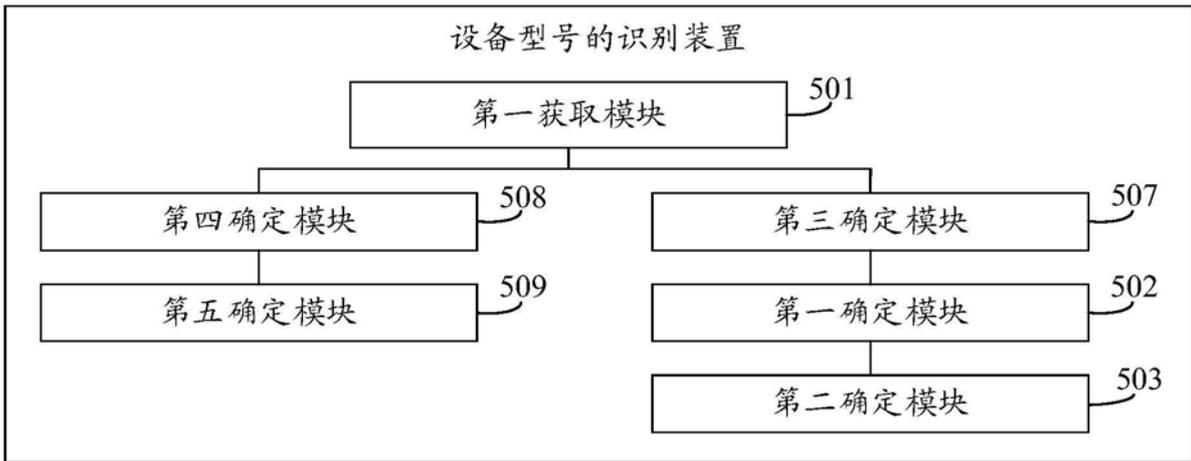


图16

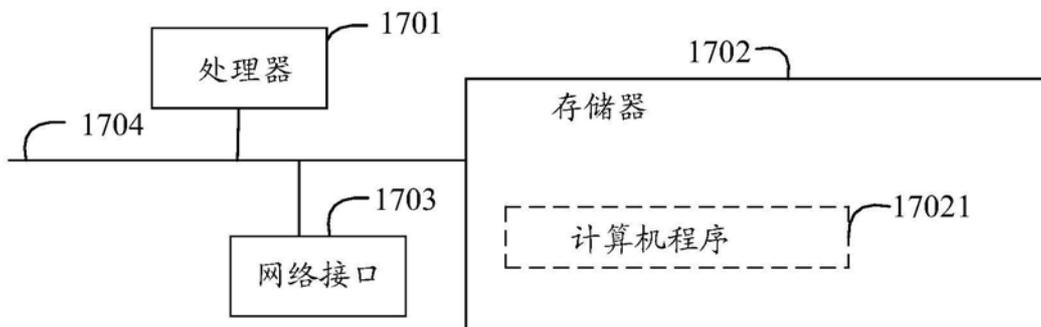


图17