(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113408552 B
(45)授权公告日 2025.02.25

(21)申请号 202010181479.9
(22)申请日 2020.03.16
(65)同一申请的已公布的文献号
申请公布号 CN 113408552 A
(43)申请公布日 2021.09.17
(73)专利权人 京东方科技集团股份有限公司
地址 100015 北京市朝阳区酒仙桥路10号
专利权人 北京大学
(72)发明人 方奕庚 穆亚东 唐小军
(74)专利代理机构 北京银龙知识产权代理有限公司 11243
专利代理师 许静 胡影
(51)Int.Cl.
G06F 18/214(2023.01)
G06F 18/25(2023.01)
G06F 16/903(2019.01)
(56)对比文件
CN 102495865 A,2012.06.13
CN 110399856 A,2019.11.01
审查员 张卓宁

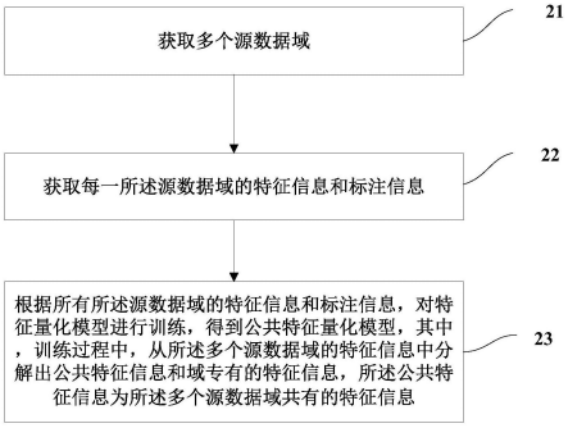
权利要求书2页 说明书9页 附图7页

(54)发明名称

特征量化模型训练、特征量化、数据查询方法及系统

(57)摘要

本发明实施例提供一种特征量化模型训练、特征量化、数据查询方法及系统，该特征量化模型训练方法包括：获取多个源数据域；获取每一所述源数据域的特征信息和标注信息；根据所有所述源数据域的特征信息和标注信息，对特征量化模型进行训练，得到公共特征量化模型，其中，训练过程中，从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息，所述公共特征信息为所述多个源数据域共有的特征信息。本发明实施例中，使用多个源数据域的丰富的标注信息训练得到公共特征量化模型，公共特征量化模型可用于标注信息匮乏的目标数据域的特征量化，从而提高特征量化模型在标注信息匮乏的数据域的特征量化性能。



CN 113408552 B

1.一种特征量化模型训练方法,其中,包括:

获取多个源数据域;所述源数据域为包括多个图像的图像数据库;

获取每一所述源数据域的特征信息和标注信息;

根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型,其中,训练过程中,从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息,所述公共特征信息为所述多个源数据域共有的特征信息;

其中,所述对特征量化模型进行训练包括:

调整所述特征量化模型,使得对于所有所述源数据域,$Ex(L(F_0(X),Y))$取最小值;

其中,X为表示所有所述源数据域的特征信息,Y为所有所述源数据域的标注信息,F_0表示公共特征量化模型,$F_0(X)$表示特征信息X经过F_0处理后得到的特征量化码,$L(F_0(X),Y)$表示所述特征量化码与标注信息Y之间的损失函数,$Ex(L(F_0(X),Y))$表示L函数针对特征信息X的数学期望;

其中,所述对特征量化模型进行训练还包括:

调整所述特征量化模型,使得对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))$取最小值,以及,对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))<Ex(L(\varphi(F_0(x),F_p(x)),y))$,其中,p不等于k;

其中,x表示所述源数据域k的特征信息,y为所述源数据域k的标注信息,F_0表示公共特征量化模型,$F_0(x)$表示特征信息x经过F_0处理后得到的特征量化码,F_k表示所述源数据域k的域专有的特征量化模型,$F_k(x)$表示特征信息x经过F_k处理后得到的特征量化码,Fp表示所述源数据域p的域专有的特征量化模型,$F_p(x)$表示特征信息x经过F_p处理后得到的特征量化码,$\varphi(F_0(x),F_k(x))$表示对$F_0(x)$和$F_k(x)$进行融合处理,$\varphi(F_0(x),F_p(x))$表示对$F_0(x)$和$F_p(x)$进行融合处理,$L(\varphi(F_0(x),F_k(x)),y)$和$L(\varphi(F_0(x),F_p(x)),y)$表示经过融合处理后的特征量化码与标注信息y之间的损失函数,Ex()表示数学期望函数,k=1,2,…,K,p=1,2,…,K,K为所述源数据域的个数。

2.如权利要求1所述的特征量化模型训练方法,其中,所述根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型包括:

根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型和每一所述源数据域的域专有的特征量化模型。

3.如权利要求2所述的特征量化模型训练方法,其中,采用深度神经网络算法对特征量化模型进行训练。

4.如权利要求1所述的特征量化模型训练方法,其中,采用相加或者线性拼接的方法进行所述融合处理。

5.一种特征量化方法,其中,包括:

采用公共特征量化模型对目标数据集进行特征量化,得到目标数据集的特征量化码,所述公共特征量化模型采用如权利要求1-4任一项所述的特征量化模型训练方法训练得到,所述目标数据集为图像数据集。

6.一种数据查询方法,其中,应用于服务器,所述方法包括:

接收客户端发送的目标查询数据的目标特征量化码;

将所述目标特征量化码与目标数据集的特征量化码进行比对,得到与所述目标特征量

化码匹配的查询结果,其中,所述目标数据集的特征量化码采用如权利要求5所述的特征量化方法得到;所述目标数据集为图像数据集;

将所述查询结果返回至所述客户端。

7.如权利要求6所述的数据查询方法,其中,所述目标数据集的特征量化码是预先采用公共特征量化模型对所述目标数据集进行特征量化得到并存储的。

8.一种数据查询方法,其中,应用于客户端,所述方法包括:

获取输入的目标查询数据;

根据公共特征量化模型,对所述目标查询数据进行特征量化计算,得到所述目标查询数据的目标特征量化码,所述公共特征量化模型采用如权利要求1-4任一项所述的特征量化模型训练方法训练得到;

将所述目标特征量化码发送给服务器;

接收所述服务器针对所述目标特征量化码返回的查询结果。

9.一种电子设备,其特征在于,包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述计算机程序被所述处理器执行时实现如权利要求1至4中任一项所述的特征量化模型训练方法的步骤。

10.一种电子设备,其特征在于,包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述计算机程序被所述处理器执行时实现如权利要求5所述的特征量化方法的步骤。

11.一种电子设备,其特征在于,包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述计算机程序被所述处理器执行时实现如权利要求6或7所述的数据查询方法的步骤。

12.一种电子设备,其特征在于,包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述计算机程序被所述处理器执行时实现如权利要求8所述的数据查询方法的步骤。

13.一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上存储计算机程序,所述计算机程序被处理器执行时实现如权利要求1至4中任一项所述的特征量化模型训练方法的步骤;或者,所述计算机程序被处理器执行时实现如权利要求5所述的特征量化方法的步骤;或者,所述计算机程序被处理器执行时实现如权利要求6或7所述的数据查询方法的步骤;或者,所述计算机程序被处理器执行时实现如权利要求8所述的数据查询方法的步骤。

特征量化模型训练、特征量化、数据查询方法及系统

技术领域

[0001] 本发明实施例涉及数据处理技术领域,尤其涉及一种特征量化模型训练、特征量化、数据查询方法及系统。

背景技术

[0002] 特征量化(feature quantization)是在计算机视觉、数据挖掘等人工智能相关领域中的一个重要技术。特征量化的目标是输出包含浓缩后的原始信息(原始的图像、视频、文本等数据的特征)的精简特征编码(特征量化码),同时能最大限度保持原始特征的表达能力。特征量化的意义在于,对于大规模数据集(如图像搜索系统中的海量图像数据),通过使用量化后的精简特征编码,能以更小的存储和计算复杂度完成特定任务(如图像搜索等)。例如,在图像搜索领域,主流的图像特征维度通常为上万维,代表性视觉特征如局部聚合描述符(VLAD)、FisherVector或者深度网络经过全局平均池化后的特征向量。在进行图像搜索等操作时,高维特征需要极高的存储代价和计算复杂度。特征量化能在基本不损失精度的情况下,极大降低对存储空间的需求和运行时刻的计算复杂度。特别的,对于百万量级的图像数据集,经过特征量化操作以后,整个数据集的特征通常只有若干吉字节(GB),可以轻易读入单台服务器的内存中,从而避免了耗时的云服务中的多机通信和内存-外存之间的输入输出(I/O)代价。

[0003] 传统的特征量化算法包括K均值聚类等。这些算法通常是无监督的,特征之间的距离或相似度计算常基于标准的欧氏距离或者余弦相似度。近年来,基于标注信息的特征量化算法逐步取得更大的关注,在实际应用中表现出更强大的性能。常见的标注信息的形式包括语义标签(例如对图像的语义类别给出一个或者多个标签)、相似度标签(例如指定两张图像是否相似、甚至具体的相似度数值)等。然而,在特定的目标数据域使用特征量化算法时,一种常见的问题是标注信息的缺乏。一方面,标注信息的获取常需要人工标注,代价昂贵;另一方面,某些垂直领域应用的标注信息在本质上是稀疏的,例如精细类别识别问题(fine-grained recognition)。从而难以保证特征量化算法的性能。

发明内容

[0004] 本发明实施例提供一种特征量化模型训练、特征量化、数据查询方法及系统,用于解决目标数据域的标注信息不足时,难以保证特征量化算法的性能的问题。

[0005] 为了解决上述技术问题,本发明是这样实现的:

[0006] 第一方面,本发明实施例提供了一种特征量化模型训练方法,包括:

[0007] 获取多个源数据域;

[0008] 获取每一所述源数据域的特征信息和标注信息;

[0009] 根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型,其中,训练过程中,从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息,所述公共特征信息为所述多个源数据域共有的特征信息。

[0010] 可选的,所述根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型包括:

[0011] 根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型和每一所述源数据域的域专有的特征量化模型。

[0012] 可选的,采用深度神经网络算法对所述公共特征量化模型和域专有的特征量化模型进行训练。

[0013] 可选的,所述对特征量化模型进行训练包括:

[0014] 调整所述特征量化模型,使得对于所有所述源数据域,$Ex(L(F_0(X),Y))$取最小值;

[0015] 其中,X为表示所有所述源数据域的特征信息,Y为所有所述源数据域的标注信息,F_0表示公共特征量化模型,$F_0(X)$表示特征信息X经过F_0处理后得到的特征量化码,$L(F_0(X),Y)$表示所述特征量化码与标注信息Y之间的损失函数,$Ex(L(F_0(X),Y))$表示L函数针对特征信息X的数学期望。

[0016] 可选的,所述对特征量化模型进行训练还包括:

[0017] 调整所述特征量化模型,使得对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))$取最小值,以及,对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))<Ex(L(\varphi(F_0(x),F_p(x)),y))$,其中,p不等于k;

[0018] 其中,x表示所述源数据域k的特征信息,y为所述源数据域k的标注信息,F_0表示公共特征量化模型,$F_0(x)$表示特征信息x经过F_0处理后得到的特征量化码,F_k表示所述源数据域k的域专有的特征量化模型,$F_k(x)$表示特征信息x经过F_k处理后得到的特征量化码,Fp表示所述源数据域p的域专有的特征量化模型,$F_p(x)$表示特征信息x经过F_p处理后得到的特征量化码,$\varphi(F_0(x),F_p(x))$表示对$F_0(x)$和$F_p(x)$进行融合处理,$\varphi(F_0(x),F_k(x))$表示对$F_0(x)$和$F_k(x)$进行融合处理,$L(\varphi(F_0(x),F_k(x)),y)$和$L(\varphi(F_0(x),F_p(x)),y)$表示经过融合处理后的特征量化码与标注信息y之间的损失函数,Ex()表示数学期望函数,k=1,2,…,K,p=1,2,…,K,K为所述源数据域的个数。

[0019] 可选的,采用相加或者线性拼接的方法进行所述融合处理。

[0020] 第二方面,本发明实施例提供了一种特征量化方法,包括:

[0021] 采用公共特征量化模型对目标数据集进行特征量化,得到目标数据集的特征量化码,所述公共特征量化模型采用上述第一方面的特征模型的信令方法训练得到。

[0022] 第三方面,本发明实施例提供了一种数据查询方法,应用于服务器,所述方法包括:

[0023] 接收客户端发送的目标查询数据的目标特征量化码;

[0024] 将所述目标特征量化码与目标数据集的特征量化码进行比对,得到与所述目标特征量化码匹配的查询结果,其中,所述目标数据集的特征量化码采用上述第二方面的特征量化方法得到;

[0025] 将所述查询结果返回至所述客户端。

[0026] 可选的,所述目标数据集的特征量化码是预先采用公共特征量化模型对所述目标数据集进行特征量化得到并存储的。

[0027] 第四方面,本发明实施例提供了一种数据查询方法,应用于客户端,所述方法包括:

[0028] 获取输入的目标查询数据；

[0029] 根据公共特征量化模型，对所述目标查询数据进行特征量化计算，得到所述目标查询数据的目标特征量化码，所述公共特征量化模型采用上述第一方面的特征量化模型训练方法训练得到；

[0030] 将所述目标特征量化码发送给服务器；

[0031] 接收所述服务器针对所述目标特征量化码返回的查询结果。

[0032] 第五方面，本发明实施例提供了一种电子设备，包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现上述第一方面的特征量化模型训练方法的步骤。

[0033] 第六方面，本发明实施例提供了一种电子设备，包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现上述第二方面的特征量化方法的步骤。

[0034] 第七方面，本发明实施例提供了一种电子设备，包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现上述第三方面的数据查询方法的步骤。

[0035] 第八方面，本发明实施例提供了一种电子设备，包括处理器、存储器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现上述第四方面的数据查询方法的步骤。

[0036] 第九方面，本发明实施例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上存储计算机程序，所述计算机程序被处理器执行时实现上述第一方面的特征量化模型训练方法的步骤；或者，所述计算机程序被处理器执行时实现上述第二方面的特征量化方法的步骤；或者，所述计算机程序被处理器执行时实现上述第三方面的数据查询方法的步骤；或者，所述计算机程序被处理器执行时实现上述第四方面的数据查询方法的步骤。

[0037] 本发明实施例中，使用多个源数据域的丰富的标注信息训练得到公共特征量化模型，公共特征量化模型可用于标注信息匮乏的目标数据域的特征量化，从而提高特征量化模型在标注信息匮乏的数据域的特征量化性能。

附图说明

[0038] 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

[0039] 图1为相关技术中的特征量化方法的示意图；

[0040] 图2为本发明实施例的特征量化模型训练方法的流程示意图；

[0041] 图3为本发明实施例中的特征量化模型训练方法的示意图；

[0042] 图4为本发明实施例的特征量化方法的流程示意图；

[0043] 图5为本发明实施例的应用于服务器端的数据查询方法的流程示意图；

[0044] 图6为本发明实施例的应用于客户端的数据查询方法的流程示意图；

[0045] 图7为本发明实施例的特征量化模型的训练系统的结构示意图；

[0046] 图8为本发明实施例的特征量化系统的结构示意图；

[0047] 图9为本发明一实施例的数据查询系统的结构示意图；

[0048] 图10为本发明另一实施例的数据查询系统的结构示意图；

[0049] 图11为本发明一实施例的电子设备的结构示意图；

[0050] 图12为本发明另一实施例的电子设备的结构示意图；

[0051] 图13为本发明又一实施例的电子设备的结构示意图；

[0052] 图14为本发明又一实施例的电子设备的结构示意图。

具体实施方式

[0053] 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作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。

[0054] 请参考图1，图1为相关技术中的特征量化方法的示意图，从图1中可以看出，相关技术中，首先需要提取数据集(或称为数据域)的特征信息(即特征提取)，然后基于数据集的标注信息对特征量化模型的关键参数进行调优，最后采用得到的特征量化模型对提取到的特征信息进行特征量化，可以看出，当标注信息匮乏时，并不能保证特征量化模型的性能。

[0055] 为解决上述问题，请参考图2，本发明实施例提供一种特征量化模型训练方法，包括：

[0056] 步骤21：获取多个源数据域；

[0057] 本发明实施例中，数据域也可以称为数据集，一个数据域包括多个数据。例如，数据域为包括多个图像的图像数据库。

[0058] 所述多个源数据域具有一定的相关度，例如存在多种相同的语义类别标签。

[0059] 步骤22：获取每一所述源数据域的特征信息和标注信息；

[0060] 所述特征信息可以根据需要设置，例如图像数据集中，特征信息可以包括图像视觉信息描述子等。

[0061] 步骤23：根据所有所述源数据域的特征信息和标注信息，对特征量化模型进行训练，得到公共特征量化模型，其中，训练过程中，从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息，所述公共特征信息为所述多个源数据域共有的特征信息。

[0062] 公共特征信息为跨域不变的公共信息，包含了多个数据域的知识。举例来说，不同摄像机的姿态不同，所拍摄到的人脸或人体的姿态也相应地存在不同，但是这些图像中存在一些共同之处，例如，人脸的拓扑结构，即人脸的拓扑结构即为公共特征信息。

[0063] 本发明实施例中，使用多个源数据域的丰富的标注信息训练得到公共特征量化模型，公共特征量化模型可用于标注信息匮乏的目标数据域的特征量化，从而提高特征量化模型在标注信息匮乏的数据域的特征量化性能。

[0064] 以面向语义检索任务的图像特征量化为例，在将特定特征量化模型施用于某目标数据域时，现有的做法是基于该目标数据域的语义标注信息对特定特征量化模型的关键参数进行调优。当语义标注信息匮乏时，现有方法并不能保证特定特征量化模型在目标数据

域的特征量化性能。本发明实施例中,借用已有的、具有丰富标注信息的多个相关源数据域,通过复用多个相关源数据域的标注信息,训练得到公共特征量化模型,采用公共特征量化模型对目标数据域进行特征量化,来达到提升特征量化模型在目标数据集上的特征量化性能的目的。

[0065] 当然,需要说明的是,本发明实施例中,数据域并不局限于图像数据集,数据域中的数据包括但不限于图像、视频、音频等数据形式。

[0066] 本发明实施例中,可选的,所述根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型包括:

[0067] 根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型和每一所述源数据域的域专有的特征量化模型。

[0068] 其中,域专有的特征信息,是指针对某一数据域的专有的特征信息。

[0069] 请参考图3,图3为本发明实施例中的特征量化模型训练方法的示意图,从图3中可以看出,用于训练特征量化模型的数据集(也称为数据域)包括K个,特征量化模型训练时,针对每一数据集,需要获取数据集的特征信息,然后,根据所有数据集的标注信息和特征信息,对特征量化模型进行训练,训练过程中,数据集的特征信息可以分解成公共特征信息和域专有的特征信息,最终得到K+1个模型,其中一个公共特征量化模型,K个域专有的特征量化模型。

[0070] 假设给定K个源数据域(数据集),记为$\langle X_k, Y_k \rangle$,其中k=1,2,…,K。其中X_k,Y_k分别表示数据集的特征信息和标注信息(通常为矩阵形式)。为了便于论述,下文中用符号x,y分别表示某数据集的特征信息和标注信息。本发明实施例中,通过机器学习的方式生成F_0,F_1,…,F_K共K+1个模型。其中,F_0为所有K个数据域所共享,F_k为第k个数据域所专有。令$F_k(x)$表示特征信息x经过Fk处理后得到的特征量化码。$\varphi(F_i(x), F_j(x))$表示对$F_i(x)$和$F_j(x)$进行融合(例如,可以进行简单的加和或者线性拼接等)。$L(F_k(x), y)$表示经过第k个模型的处理后,特征信息x经过Fk处理后得到的特征量化码与标注信息y之间的损失函数(例如,L可以为分类0-1损失函数),我们希望获得更小的损失函数值。$Ex(L(F_k(x), y))$表示L函数针对x的数学期望。

[0071] 为了得到上述各个模型,需要对所有的源数据域$\langle X_k, Y_k \rangle$进行模型学习的过程,学习过程中的具体的优化目标包括:

[0072] 1)对于所有的k=1,2,…,K,$Ex(L(F_0(x), y))$应当取得最小值。这样保证了公共特征量化模型获得优异的特征量化性能;

[0073] 2)对于任意k=1,2,…,K,$Ex(L(\varphi(F_0(x), F_k(x)), y))$应当取得最小值。这样保证了域专有的特征量化模型与公共特征量化模型的互补性;

[0074] 3)对于任意k=1,2,…,K,$Ex(L(\varphi(F_0(x), F_k(x)), y)) < Ex(L(\varphi(F_0(x), F_p(x)), y))$,其中p不等于k。这样保证了域专有的特征量化模型对于特定数据域的最优性。

[0075] 即,所述对特征量化模型进行训练包括:

[0076] 调整所述特征量化模型,使得对于所有所述源数据域,$Ex(L(F_0(X), Y))$取最小值;

[0077] 其中,X为表示所有所述源数据域的特征信息,Y为所有所述源数据域的标注信息,F_0表示公共特征量化模型,$F_0(X)$表示特征信息X经过F_0处理后得到的特征量化码,$L(F_0(X), Y)$表示经过F_0处理后特征信息X得到的特征量化码与标注信息Y之间的损失函数,$Ex(L(F_0$

(X),Y))表示L函数针对特征信息X的数学期望。

[0078] 进一步的,所述对特征量化模型进行训练还包括:

[0079] 调整所述特征量化模型,使得对于任意所述源数据域k,Ex(L(φ(F_0(x),F_k(x)),y))取最小值;以及,对于任意所述源数据域k,Ex(L(φ(F_0(x),F_k(x)),y))<Ex(L(φ(F_0(x),F_p(x)),y)),其中,p不等于k;

[0080] 其中,x表示所述源数据域k的特征信息,y为所述源数据域k的标注信息,F_0表示公共特征量化模型,F_0(x)表示特征信息x经过F_0处理后得到的特征量化码,F_k表示所述源数据域k的域专有的特征量化模型,F_k(x)表示特征信息x经过F_k处理后得到的特征量化码,Fp表示所述源数据域p的域专有的特征量化模型,F_p(x)表示特征信息x经过F_p处理后得到的特征量化码,φ(F_0(x),F_k(x))表示对F_0(x)和F_k(x)进行融合处理,φ(F_0(x),F_p(x))表示对F_0(x)和F_p(x)进行融合处理,L(φ(F_0(x),F_k(x)),y)和L(φ(F_0(x),F_p(x)),y)表示经过融合处理后的特征量化码与标注信息y之间的损失函数,Ex()表示数学期望函数,k=1,2,…,K,p=1,2,…,K,K为所述源数据域的个数。

[0081] 本发明实施例中,可选的,采用相加或者线性拼接的方法进行所述融合处理。

[0082] 本发明实施例中,对于不同的源数据域,将域专有的特征量化模型与公共特征量化模型的结果进行融合后,相对仅仅使用公共特征量化模型,能保证提升在该数据域的特征量化性能。

[0083] 本发明实施例中,对于不同的数据域,还可以交换使用彼此的域专有的特征量化模型,并与公共特征量化模型融合,其实际效果将约等于引入随机噪声,或者引起严重的过拟合现象。

[0084] 本发明实施例中,可选的,采用深度神经网络算法对特征量化模型进行训练。例如,可基于多层卷积、池化或非线性激活网络层对特征量化模型进行训练。

[0085] 本发明实施例中,可以采用多种方式提取每一所述源数据域的特征信息,例如可以采用深度神经网络算法提取每一所述源数据域的特征信息。

[0086] 本发明实施例中,可选的,公共特征量化模型和域专有的特征量化模型采用局部敏感哈希算法或者K均值算法。进一步可选的,若数据集为图像数据集,公共特征量化模型和域专有的特征量化模型采用局部敏感哈希算法。

[0087] 本发明实施例中,可选的,若数据集为图像数据集,针对图像检索任务,可采用以下方式:1)图像特征提取基于预训练神经网络(如ResNet50等);2)公共特征量化模型和域专有的特征量化模型采取浅层卷积网络;3)公共特征量化模型和域专有的特征量化模型采取线性拼接方式融合。

[0088] 本发明实施例中,上述特征量化模型训练方法可以由服务器端执行。

[0089] 请参考图4,本发明实施例还提供一种特征量化方法,包括:

[0090] 步骤41:采用公共特征量化模型对目标数据集进行特征量化,得到目标数据集的特征量化码,所述公共特征量化模型采用上述特征量化模型训练方法训练得到。

[0091] 本发明实施例中,使用多个源数据域的丰富的标注信息训练得到公共特征量化模型,公共特征量化模型可用于标注信息匮乏的目标数据域的特征量化,从而提高特征量化模型在标注信息匮乏的数据域的特征量化性能。

[0092] 请参考图5,本发明实施例还提供一种数据查询方法,所述数据查询方法应用于服

务器端,包括:

[0093] 步骤51:接收客户端发送的目标查询数据的目标特征量化码;

[0094] 步骤52:将所述目标特征量化码与目标数据集的特征量化码进行比对,得到与所述目标特征量化码匹配的查询结果,其中,所述目标数据集的特征量化码采用上述特征量化方法得到;

[0095] 步骤53:将所述查询结果返回至所述客户端。

[0096] 可选的,所述目标数据集的特征量化码是预先采用公共特征量化模型对所述目标数据集进行特征量化得到并存储的。

[0097] 请参考图6,本发明实施例还提供一种数据查询方法,所述数据查询方法应用于客户端,包括:

[0098] 步骤61:获取输入的目标查询数据;

[0099] 步骤62:根据公共特征量化模型,对所述目标查询数据进行特征量化计算,得到所述目标查询数据的目标特征量化码,所述公共特征量化模型采用上述特征量化模型训练方法训练得到。

[0100] 请参考图7,本发明实施例还提供一种特征量化模型的训练系统70,包括:

[0101] 第一获取模块71,用于获取多个源数据域;

[0102] 第二获取模块72,用于获取每一所述源数据域的特征信息和标注信息;

[0103] 训练模块73,用于根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型,其中,训练过程中,从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息,所述公共特征信息为所述多个源数据域共有的特征信息。

[0104] 可选的,所述训练模块73,用于根据所有所述源数据域的特征信息和标注信息,对特征量化模型进行训练,得到公共特征量化模型和每一所述源数据域的域专有的特征量化模型。

[0105] 可选的,所述训练模块73,用于采用深度神经网络算法对特征量化模型进行训练。

[0106] 可选的,所述训练模块73,用于调整所述特征量化模型,使得对于所有所述源数据域,$Ex(L(F_0(X),Y))$取最小值;

[0107] 其中,X为表示所有所述源数据域的特征信息,Y为所有所述源数据域的标注信息,F_0表示公共特征量化模型,$F_0(X)$表示特征信息X经过F_0处理后得到的特征量化码,$L(F_0(X),Y)$表示所述特征量化码与标注信息Y之间的损失函数,$Ex(L(F_0(X),Y))$表示L函数针对特征信息X的数学期望。

[0108] 可选的,所述训练模块73,用于调整所述特征量化模型,使得对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))$取最小值,以及,对于任意所述源数据域k,$Ex(L(\varphi(F_0(x),F_k(x)),y))<Ex(L(\varphi(F_0(x),F_p(x)),y))$,其中,p不等于k;

[0109] 其中,x表示所述源数据域k的特征信息,y为所述源数据域k的标注信息,F_0表示公共特征量化模型,$F_0(x)$表示特征信息x经过F_0处理后得到的特征量化码,F_k表示所述源数据域k的域专有的特征量化模型,$F_k(x)$表示特征信息x经过F_k处理后得到的特征量化码,Fp表示所述源数据域p的域专有的特征量化模型,$F_p(x)$表示特征信息x经过F_p处理后得到的特征量化码,$\varphi(F_0(x),F_k(x))$表示对$F_0(x)$和$F_k(x)$进行融合处理,$\varphi(F_0(x),F_p(x))$表示对$F_0(x)$

和$F_p(x)$进行融合处理,$L(\varphi(F_0(x),F_k(x)),y)$和$L(\varphi(F_0(x),F_p(x)),y)$表示经过融合处理后的特征量化码与标注信息y之间的损失函数,Ex()表示数学期望函数,k=1,2,…,K,p=1,2,…,K,K为所述源数据域的个数。

[0110] 可选的,所述训练模块73,用于采用相加或者线性拼接的方法进行所述融合处理。

[0111] 请参考图8,本发明实施例还提供一种特征量化系统80,包括:

[0112] 特征量化模块81,用于采用公共特征量化模型对目标数据集进行特征量化,得到目标数据集的特征量化码,所述公共特征量化模型采用上述特征量化模型训练方法训练得到。

[0113] 所述特征量化系统80可以为服务器。

[0114] 请参考图9,本发明实施例还提供一种数据查询系统90,包括:

[0115] 接收模块91,用于接收客户端发送的目标查询数据的目标特征量化码;

[0116] 查询模块92,用于将所述目标特征量化码与目标数据集的特征量化码进行比对,得到与所述目标特征量化码匹配的查询结果,其中,所述目标数据集的特征量化码采用上述特征量化方法得到;

[0117] 发送模块93,用于将所述查询结果返回至所述客户端。

[0118] 所述数据查询系统90可以为服务器。

[0119] 可选的,所述目标数据集的特征量化码是预先采用公共特征量化模型对所述目标数据集进行特征量化得到并存储的。

[0120] 请参考图10,本发明实施例还提供一种数据查询系统100,包括:

[0121] 获取模块101,用于获取输入的目标查询数据;

[0122] 计算模块102,用于根据公共特征量化模型,对所述目标查询数据进行特征量化计算,得到所述目标查询数据的目标特征量化码,所述公共特征量化模型采用上述特征量化模型训练方法训练得到;

[0123] 发送模块103,用于将所述目标特征量化码发送给服务器;

[0124] 接收模块104,用于接收所述服务器针对所述目标特征量化码返回的查询结果。

[0125] 所述数据查询系统100可以为客户端。

[0126] 请参考图11,本发明实施例还提供一种电子设备110,包括处理器111,存储器112,存储在存储器112上并可在所述处理器111上运行的计算机程序,该计算机程序被处理器111执行时实现上述特征量化模型训练方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0127] 可选的,所述电子设备110为服务器。

[0128] 请参考图12,本发明实施例还提供一种电子设备120,包括处理器121,存储器122,存储在存储器122上并可在所述处理器121上运行的计算机程序,该计算机程序被处理器121执行时实现上述特征量化方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0129] 可选的,所述电子设备120为服务器。

[0130] 请参考图13,本发明实施例还提供一种电子设备130,包括处理器131,存储器132,存储在存储器132上并可在所述处理器131上运行的计算机程序,该计算机程序被处理器131执行时实现上述应用于服务器的数据查询方法实施例的各个过程,且能达到相同的技

术效果,为避免重复,这里不再赘述。

[0131] 可选的,所述电子设备130为服务器。

[0132] 请参考图14,本发明实施例还提供一种电子设备140,包括处理器141,存储器142,存储在存储器142上并可在所述处理器141上运行的计算机程序,该计算机程序被处理器141执行时实现上述应用于客户端的数据查询方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0133] 可选的,所述电子设备140为客户端。

[0134] 本发明实施例还提供一种计算机可读存储介质,计算机可读存储介质上存储有计算机程序,该计算机程序被处理器执行时实现上述特征量化模型训练方法实施例的各个过程,或者,该计算机程序被处理器执行时实现上述特征量化方法实施例的各个过程,或者,该计算机程序被处理器执行时实现上述应用于服务器端的数据查询方法实施例的各个过程,或者,该计算机程序被处理器执行时实现上述应用于客户端的数据查询方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。其中,所述的计算机可读存储介质,如只读存储器(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等。

[0135] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0136] 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上述实施例方法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,但很多情况下前者是更佳的实施方式。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(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0137] 上面结合附图对本发明的实施例进行了描述,但是本发明并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本发明的启示下,在不脱离本发明宗旨和权利要求所保护的范围情况下,还可做出很多形式,均属于本发明的保护之内。

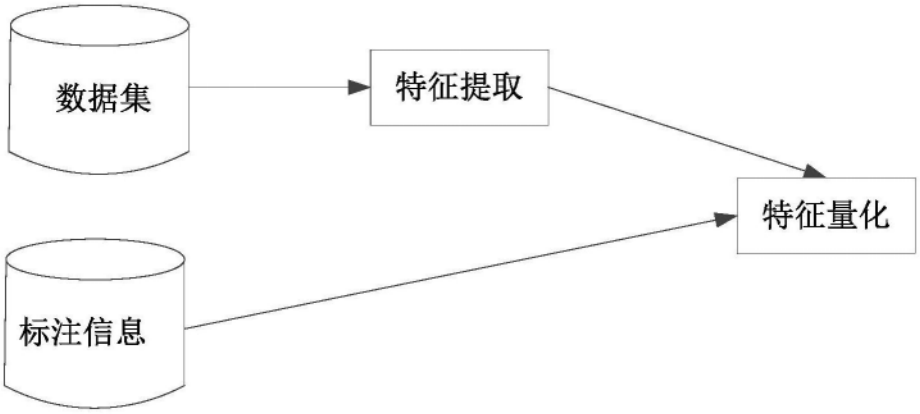

数据集
特征提取
特征量化
标注信息

图1

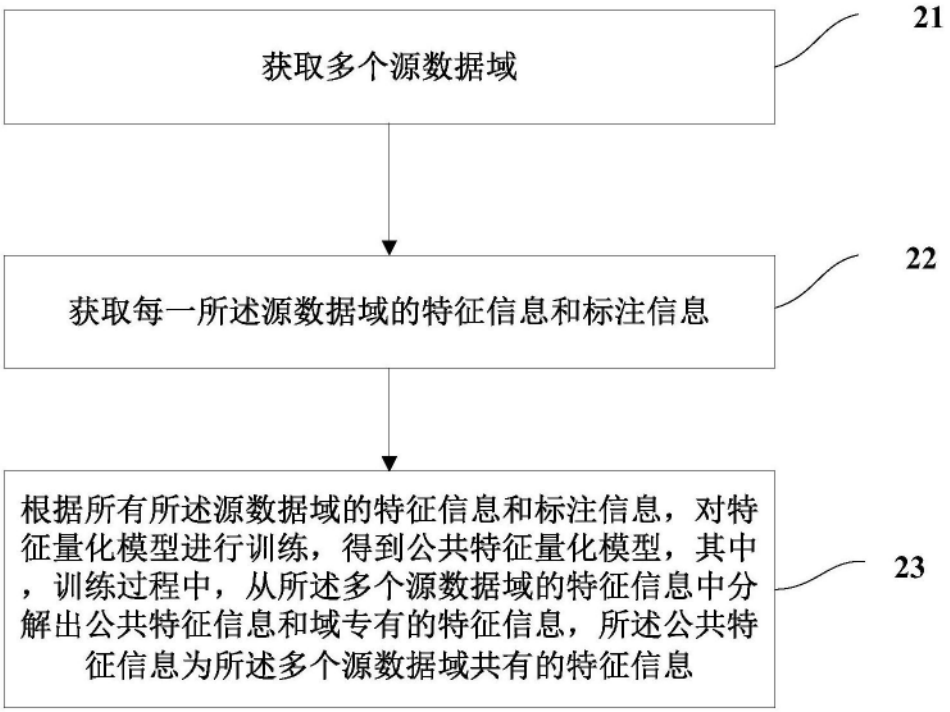

获取多个源数据域
21
获取每一所述源数据域的特征信息和标注信息
22
根据所有所述源数据域的特征信息和标注信息，对特征量化模型进行训练，得到公共特征量化模型，其中，训练过程中，从所述多个源数据域的特征信息中分解出公共特征信息和域专有的特征信息，所述公共特征信息为所述多个源数据域共有的特征信息
23

图2

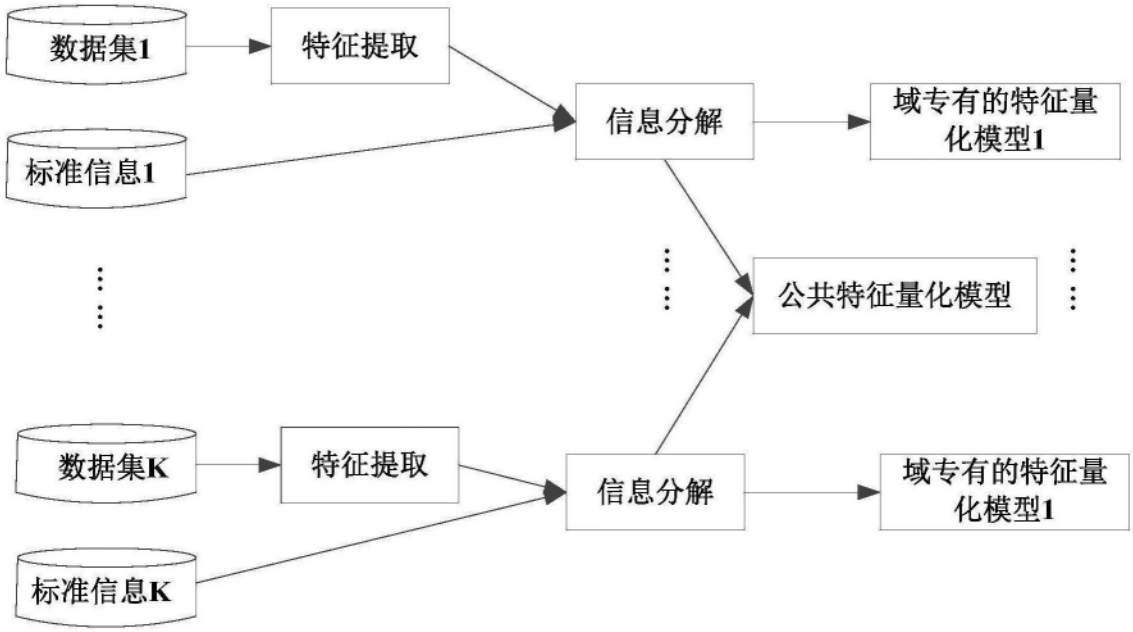

数据集1
特征提取
标准信息1
信息分解
域专有的特征量化模型1
公共特征量化模型
数据集K
特征提取
标准信息K
信息分解
域专有的特征量化模型1

图3

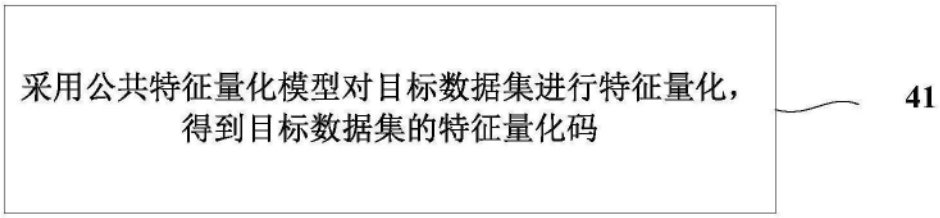

采用公共特征量化模型对目标数据集进行特征量化，得到目标数据集的特征量化码
41

图4

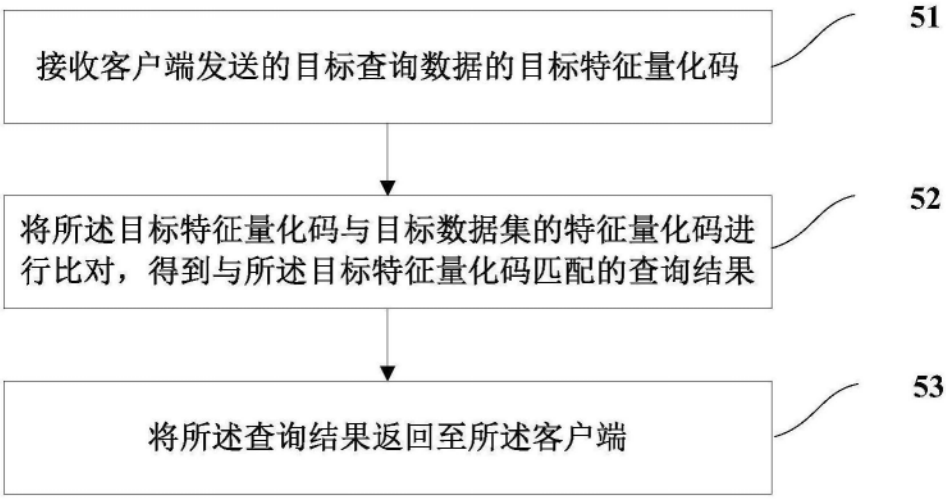

接收客户端发送的目标查询数据的目标特征量化码
51
将所述目标特征量化码与目标数据集的特征量化码进行比对，得到与所述目标特征量化码匹配的查询结果
52
将所述查询结果返回至所述客户端
53

图5

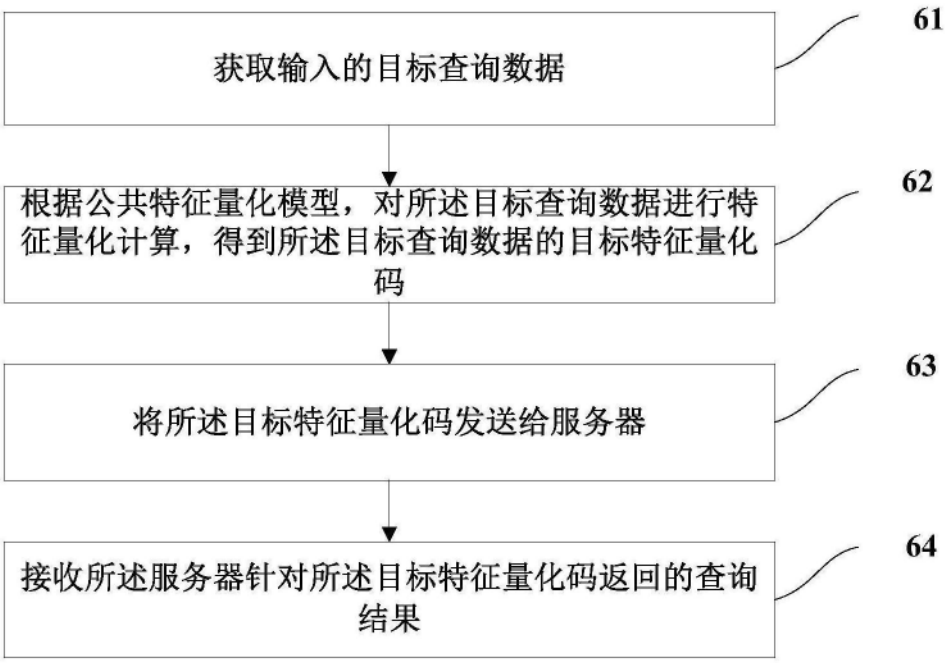

获取输入的目标查询数据
61
根据公共特征量化模型，对所述目标查询数据进行特征量化计算，得到所述目标查询数据的目标特征量化码
62
将所述目标特征量化码发送给服务器
63
接收所述服务器针对所述目标特征量化码返回的查询结果
64

图6

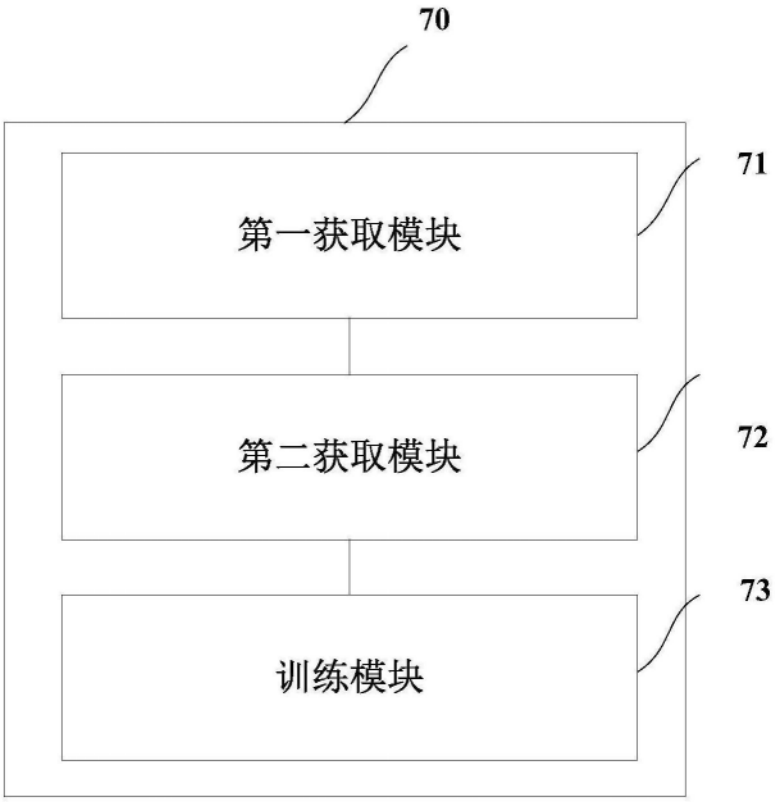

70
第一获取模块
71
第二获取模块
72
训练模块
73

图7

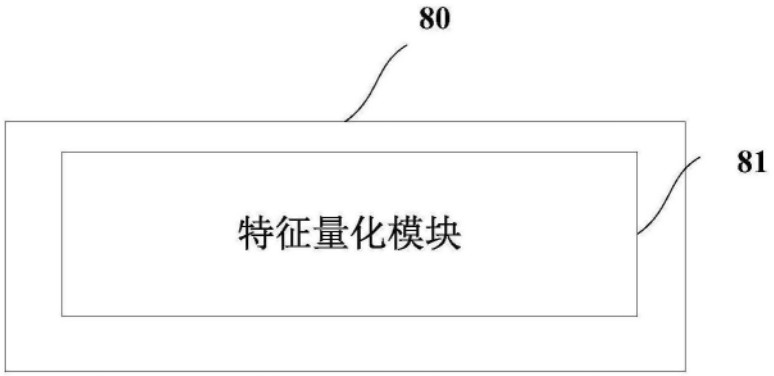


图8

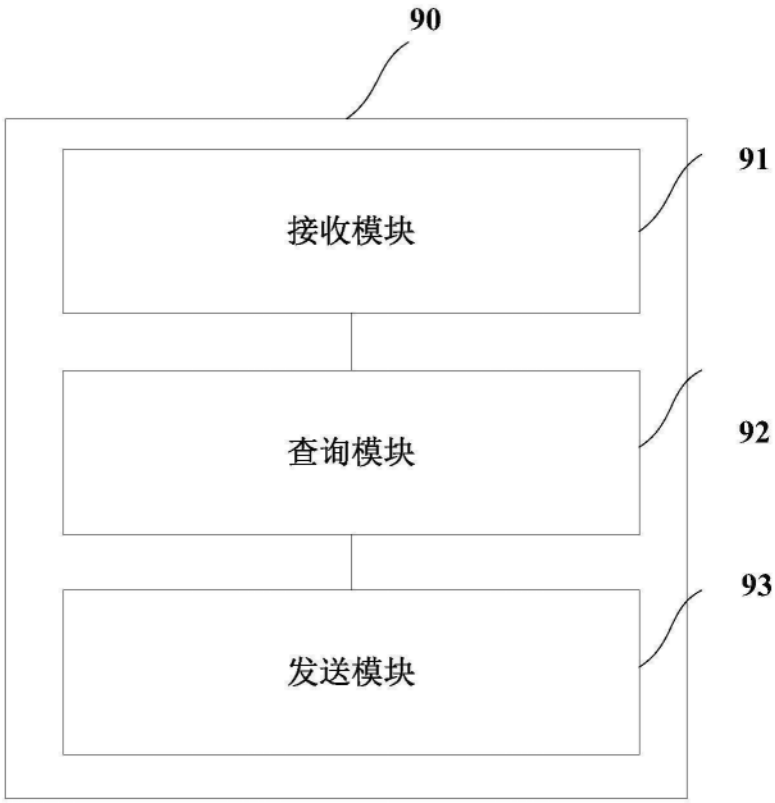


图9

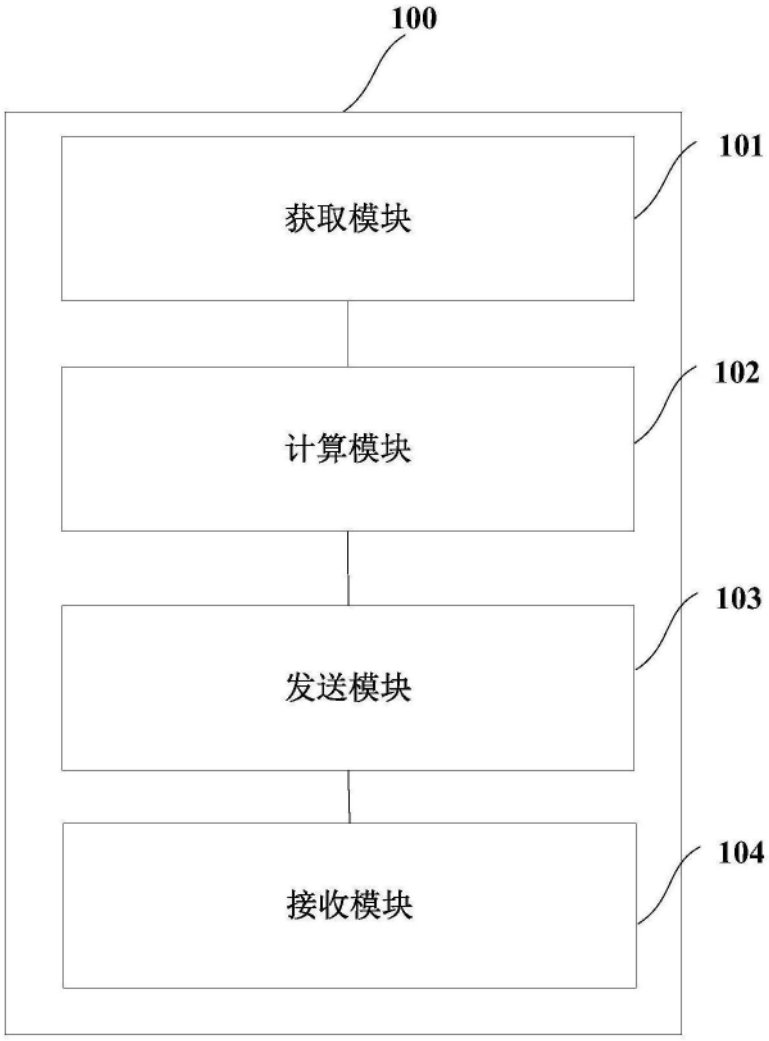


图10

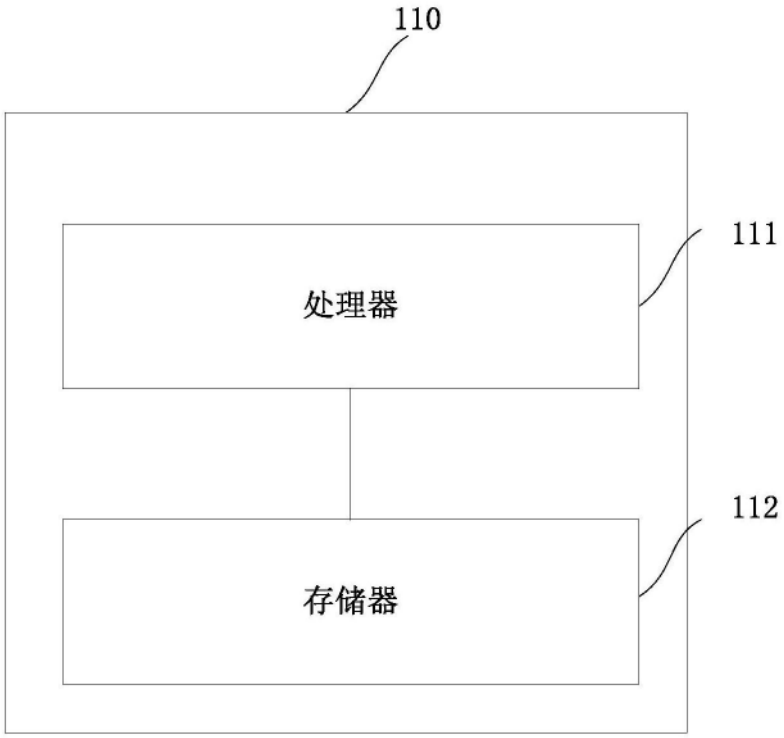


图11

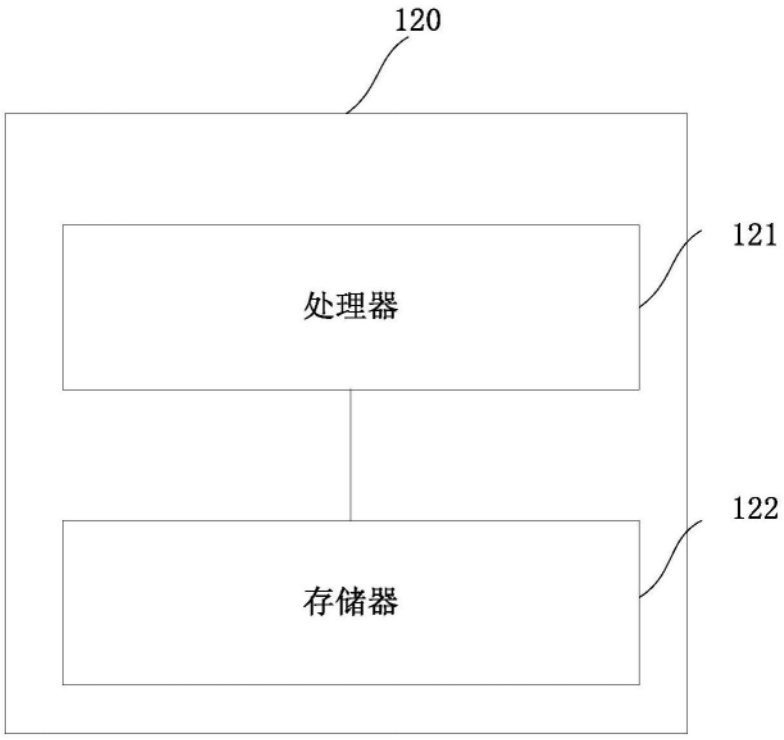


图12

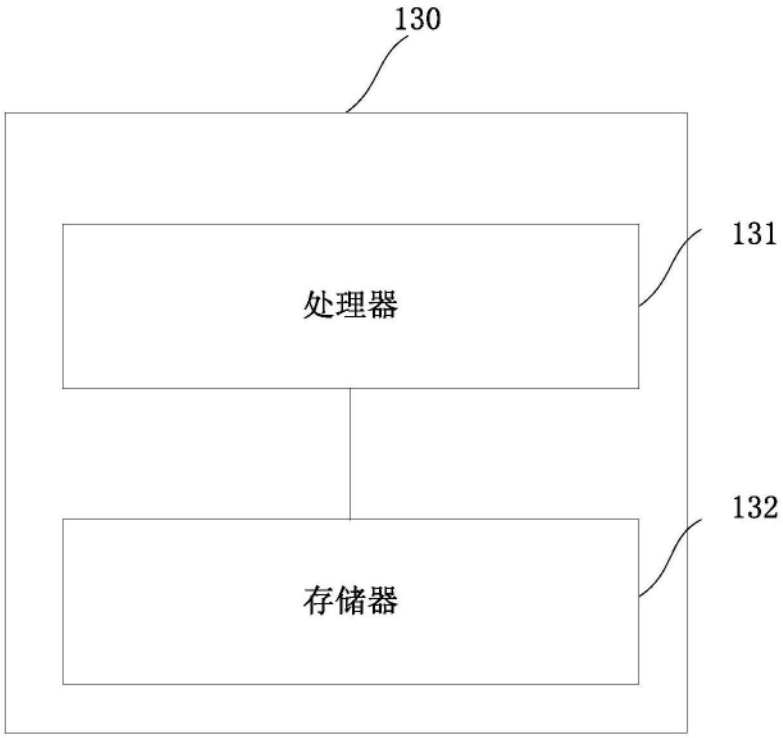


图13

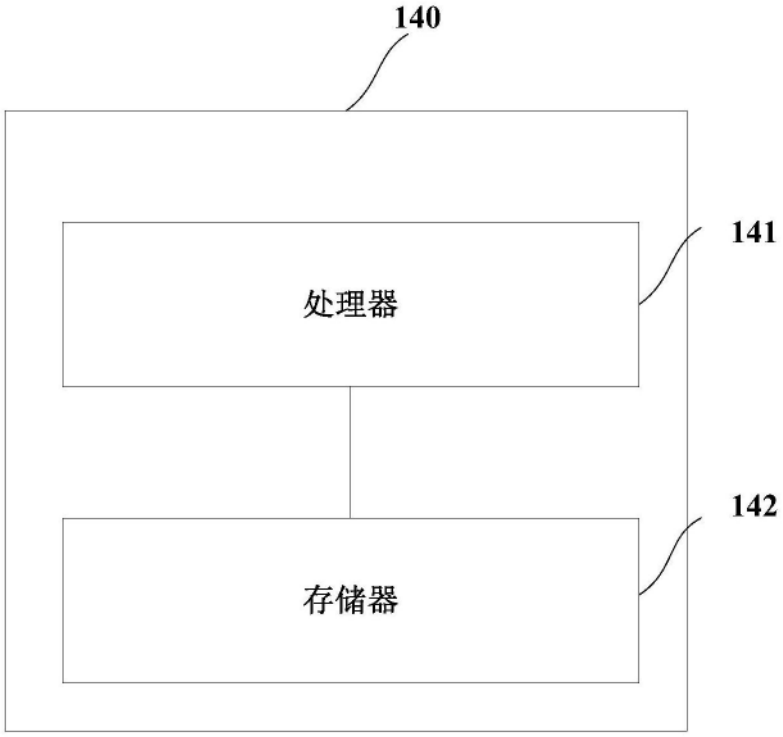


图14