(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 112287208 A
(43)申请公布日 2021.01.29

(21)申请号 201910940066.1

(22)申请日 2019.09.30

(71)申请人 北京沃东天骏信息技术有限公司
地址 100176 北京市大兴区经济技术开发区科创十一街18号院2号楼4层A402室

(72)发明人 余鑫 王蒙 王发庆 于亚男 阚景森

(74)专利代理机构 北京律智知识产权代理有限公司 11438
代理人 王辉 阚梓瑄

(51)Int.Cl.
G06F 16/9535(2019.01)
G06F 16/18(2019.01)

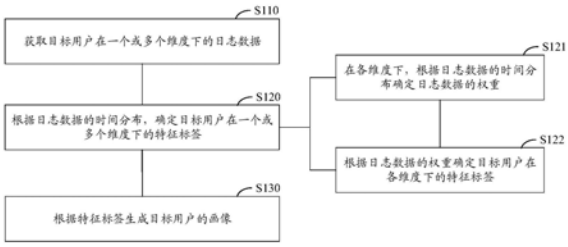
权利要求书2页 说明书11页 附图3页

(54)发明名称

用户画像生成方法、装置、电子设备及存储介质

(57)摘要

本申请提供了一种用户画像生成方法、装置、电子设备及计算机可读存储介质。该方法包括:获取目标用户在一个或多个维度下的日志数据;根据日志数据的时间分布,确定目标用户在所述一个或多个维度下的特征标签;根据特征标签生成所述目标用户的画像;其中,所述根据所述日志数据的时间分布,确定所述目标用户在所述一个或多个维度下的特征标签包括:在各所述维度下,根据所述日志数据的时间分布确定所述日志数据中各数据的权重;根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签。本申请可以提高用户画像生成的准确性,从而实现生成的用户画像较为全面。



CN 112287208 A

1.一种用户画像生成方法,其特征在于,包括:

获取目标用户在一个或多个维度下的日志数据;

根据所述日志数据的时间分布,确定所述目标用户在所述一个或多个维度下的特征标签;

根据所述特征标签生成所述目标用户的画像;

其中,所述根据所述日志数据的时间分布,确定所述目标用户在所述一个或多个维度下的特征标签包括:

在各所述维度下,根据所述日志数据的时间分布确定所述日志数据中各数据的权重;

根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签。

2.根据权利要求1所述的方法,其特征在于,所述根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签包括:

在各所述维度下,根据所述日志数据的各数据中权重最大的数据确定所述目标用户在所述维度下的特征标签。

3.根据权利要求1所述的方法,其特征在于,所述在各所述维度下,根据所述日志数据的时间分布确定所述日志数据中各数据的权重包括:

在各所述维度下,按照预设的周期对所述日志数据进行统计,以得到所述日志数据中各数据所对应的周期序数;

根据所述日志数据中各数据所对应的周期序数确定所述各数据的权重。

4.根据权利要求3所述的方法,其特征在于,所述根据所述日志数据中各数据所对应的预设周期的序数确定所述日志数据的权重包括:

基于所述日志数据中各数据所对应的周期序数,通过指数函数确定所述各数据的权重;其中,所述周期序数为所述指数函数的指数,所述指数函数的底数为常数。

5.根据权利要求3所述的方法,其特征在于,所述方法还包括:

确定所述日志数据中各数据的出现频次;

所述根据所述日志数据中各数据所对应的周期序数确定所述各数据的权重包括:

对于所述日志数据中任一数据D_i,通过以下公式计算数据D_i的权重:

$$B(D_i) = B(S_{i1}) + B(S_{i2}) + \cdots + B(S_{im}) = freq(D_i) \cdot [e^{-S_{i1}{}^k+1} + e^{-S_{i2}{}^k+1} + \cdots + e^{-S_{im}{}^k+1}];$$

其中,B表示权重,S_{i1}、S_{i2}、…、S_{im}为数据D_i对应的周期序数,$freq(D_i)$为数据D_i在所述日志数据中的出现频次,k为指数常数。

6.根据权利要求1所述的方法,其特征在于,所述获取目标用户在一个或多个维度下的日志数据包括:

获取所述目标用户在所述一个或多个维度下,且在预设时间范围内的日志数据。

7.一种用户画像生成装置,其特征在于,包括:

数据获取模块,用于获取目标用户在一个或多个维度下的日志数据;

标签确定模块,用于根据所述日志数据的时间分布,确定所述目标用户在所述一个或多个维度下的特征标签。

画像生成模块,用于根据所述特征标签生成所述目标用户的画像;

其中,标签确定模块包括:

权重确定单元,用于在各所述维度下,根据所述日志数据的时间分布确定所述日志数据中各数据的权重;

标签处理单元,用于根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签。

8.一种电子设备,其特征在于,包括:

处理器;以及

存储器,用于存储所述处理器的可执行指令;

其中,所述处理器配置为经由执行所述可执行指令来执行权利要求1-6任一项所述的方法。

9.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1-6任一项所述的方法。

用户画像生成方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及计算机技术领域，尤其涉及一种用户画像生成方法、装置、电子设备及计算机可读存储介质。

背景技术

[0002] 随着互联网在各个行业内越来越广泛的普及与应用，电商、互联网金融、生活服务、游戏等多个领域的企业都致力于通过互联网收集与分析用户的静态属性、社会属性、行为属性等信息数据，以抽象出用户画像，从而挖掘用户需求，为用户提供更加具有针对性的产品或服务。

[0003] 现有的用户画像生成方法大多是将用户的常用信息进行组合，例如统计出用户的常用支付方式、常用下单设备或常用收获地址等，分别作为用户的特征标签，组合为用户的画像。然而，该方法通过常用信息表征用户的特征，常用信息在一些情况下无法代表用户的当前状态，导致生成的用户画像较为片面，缺乏客观性；另外，该方法仅通过统计常用信息的方式对用户信息进行处理，其方式较为单一，灵活性较差，且无法挖掘用户信息的变化规律，导致生成的用户画像过于表面化，其准确性较低。

[0004] 需要说明的是，在上述背景技术部分公开的信息仅用于加强对本申请的背景的理解，因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

[0005] 本申请提供了一种用户画像生成方法、装置、电子设备及计算机可读存储介质，克服了现有的用户画像生成方法所生成的用户画像的准确性较差的问题。

[0006] 本申请的其他特性和优点将通过下面的详细描述变得显然，或部分地通过本申请的实践而习得。

[0007] 根据本申请的一个方面，提供用户画像生成方法，包括：获取目标用户在一个或多个维度下的日志数据；根据所述日志数据的时间分布，确定所述目标用户在所述一个或多个维度下的特征标签；根据所述特征标签生成所述目标用户的画像。

[0008] 在本申请的一种示例性实施例中，所述根据所述日志数据的时间分布，确定所述目标用户在所述一个或多个维度下的特征标签包括：在各所述维度下，根据所述日志数据的时间分布确定所述日志数据中各数据的权重；根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签；其中，所述根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签包括：在各所述维度下，根据所述日志数据的各数据中权重最大的数据确定所述目标用户在所述维度下的特征标签。

[0009] 在本申请的一种示例性实施例中，所述在各所述维度下，根据所述日志数据的时间分布确定所述日志数据中各数据的权重包括：在各所述维度下，按照预设的周期对所述日志数据进行统计，以得到所述日志数据中各数据所对应的周期序数；根据所述日志数据中各数据所对应的周期序数确定所述各数据的权重。

[0010] 在本申请的一种示例性实施例中，所述根据所述日志数据中各数据所对应的预设周期的序数确定所述日志数据的权重包括：基于所述日志数据中各数据所对应的周期序数，通过指数函数确定所述各数据的权重；其中，所述周期序数为所述指数函数的指数，所述指数函数的底数为常数。

[0011] 在本申请的一种示例性实施例中，所述方法还包括：确定所述日志数据中各数据的出现频次；所述根据所述日志数据中各数据所对应的周期序数确定所述各数据的权重包括：对于所述日志数据中任一数据D_i，通过以下公式计算数据D_i的权重：$B(D_i) = B(S_{i1}) + B(S_{i2}) + \cdots + B(S_{im}) = freq(D_i) \cdot [e^{-S_{i1}{}^k+1} + e^{-S_{i2}{}^k+1} + \cdots + e^{-S_{im}{}^k+1}]$；其中，B表示权重，$S_{i1}$、$S_{i2}$、…、$S_{im}$为数据$D_i$对应的周期序数，$freq(D_i)$为数据$D_i$在所述日志数据中的出现频次，k为指数常数。

[0012] 在本申请的一种示例性实施例中，所述获取目标用户在一个或多个维度下的日志数据包括：获取所述目标用户在所述一个或多个维度下且在预设时间范围内的日志数据。

[0013] 根据本申请的一个方面，提供一种用户画像生成装置，包括：根据所述日志数据的时间分布，确定所述目标用户在所述一个或多个维度下各维度对应的多个特征标签；所述根据所述特征标签生成所述目标用户的画像包括：根据各维度对应的所述多个特征标签生成所述目标用户的画像；其中，标签确定模块包括：权重确定单元，用于在各所述维度下，根据所述日志数据的时间分布确定所述日志数据中各数据的权重；标签处理单元，用于根据所述日志数据中各数据的权重确定所述目标用户在各所述维度下的特征标签。

[0014] 根据本申请的一个方面，提供一种电子设备，包括：处理器；以及存储器，用于存储所述处理器的可执行指令；其中，所述处理器配置为经由执行所述可执行指令来执行上述任意一项所述的方法。

[0015] 根据本申请的一个方面，提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述任意一项所述的方法。

[0016] 本申请的示例性实施例具有以下有益效果：

[0017] 通过获取目标用户的日志数据，并根据日志数据的时间分布确定特征标签，从而生成目标用户的画像。一方面，在本示例性实施例中，因为各数据所对应的时间分布可能不同，根据日志数据的时间分布确定特征标签，可以使生成的特征标签客观地反映出不同时间分布下用户特征数据的差异，生成较为全面的用户画像；另一方面，结合时间分布确定用户画像，可以使用户画像在生成时所考虑的因素更加丰富，且根据时间分布，能够有效确定目标用户的日志数据中各数据的变化规律，从而确定最接近当前状态的特征标签，提高用户画像生成的准确性。

[0018] 应当理解的是，以上的一般描述和后文的细节述仅是示例性和解释性的，并不能限制本申请。

附图说明

[0019] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。显而易见地，下面描述中的附图仅仅是本申请的一些实施例，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。

[0020] 图1示意性示出本示例性实施例中一种用户画像生成方法的流程图；

[0021] 图2示意性示出本示例性实施例中一种用户画像生成方法的子流程图；

[0022] 图3示意性示出本示例性实施例中另一种用户画像生成方法的流程图；

[0023] 图4示意性示出本示例性实施例中一种用户画像生成装置的结构框图；

[0024] 图5示意性示出本示例性实施例中一种用于实现上述方法的电子设备；

[0025] 图6示意性示出本示例性实施例中一种用于实现上述方法的计算机可读存储介质。

具体实施方式

[0026] 现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本申请将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。

[0027] 本申请的示例性实施例提供了一种用户画像生成方法，其中用户画像是指对用户信息进行标签化，通过收集与分析用户的多种信息数据，抽象出用户全貌，即用户画像。用户画像可以用于个性化推荐等大数据应用中，其应用场景可以用在包括金融服务、电商、生活服务、游戏、社交网络、音乐等各个类型的应用程序，本申请对此不做特别限定。

[0028] 下面结合附图1对本示例性实施例做进一步说明，如图1所示，用户画像生成方法可以包括以下步骤S110～S130：

[0029] 步骤S110，获取目标用户在一个或多个维度下的日志数据。

[0030] 其中，目标用户可以是需要进行画像的用户，日志数据可以是关于目标用户的原始信息数据，例如目标用户的的年龄、地址、职业、支付方式或兴趣爱好等。维度是指日志数据中包含或反映的目标用户信息的类别，通常一个维度是反映目标用户属性的一个方面，例如，维度可以包括个人信息属性、信用属性、消费特征属性或社交信息属性等。在本示例性实施例中，服务器可以实时获取用户终端的日志数据，也可以从特定的数据库中获取日志数据，例如HDFS(Hadoop Distributed File System，分布式文件系统)。在日志数据中，目标用户通常是通过唯一性标识来表征的，例如用户的手机号、App(Application，应用程序)账号、IP地址(Internet Protocol，互联网协议地址)等。举例而言，用户A登录App后，为了生成用户A的用户画像，可以识别出用户A的App账号，并根据该账号在日志数据库中搜索相应的日志数据。在获取日志数据时，可以获取全部的日志数据，以便于全面统计目标用户的信息，也可以获取某一时间范围内的日志数据，以减小后续需要处理的数据量。

[0031] 步骤S120，根据日志数据的时间分布，确定目标用户在一个或多个维度下的特征标签。

[0032] 其中，日志数据的时间分布是指日志数据在不同时间节点上的分布特征，可以表现为日志数据的时间-数值变化趋势，也可以表现为日志数据中各数据的出现频次与时效性等。本示例性实施例中，出现频次是指各数据在全部的日志数据中出现的频繁程度，时效性是指各数据距离当前时间的远近，越靠近当前时间表示该数据的时效性越强，通常出现频次越高、时效性越强的数据越能够表征目标用户当前的状态。特征标签是指在每个维度

下,最能够表现目标用户真实信息或当前状态的数据信息,通常是对目标用户的特征进行的抽象概括。特征标签可以是日志数据中的原始数据,例如目标用户的收货地址“xx市xx区xx路x号”,也可以是日志数据的关键词或摘要信息,例如“xx区”或“xx路”,还可以是基于日志数据而计算得到的信息数据,例如用户的月平均消费额等等。

[0033] 在本示例性实施例中,可以根据日志数据的时间分布,通过预设的处理方式对日志数据进行处理。例如可以将日志数据按照时间顺序转换为特征向量,将其输入预先训练的LSTM(Long Short-Term Memory,长短时记忆网络),输出相应的特征标签;也可以将日志数据绘制为以时间、数值为坐标的图线,并进行函数拟合,根据拟合的结果确定相应的特征标签;还可以利用预设的计算公式对日志数据进行计算,得到相应的特征标签等。

[0034] 在一示例性实施例中,可以根据目标用户的日志数据的时间分布,确定时间粒度并对日志数据进行划分,划分后的日志数据属于各自对应的时间粒度,例如以月为时间粒度,可以根据日志数据的时间分布将各日志数据划分至相应的月度中,使日志数据的时间分布表现为数据在不同月度中的分布特征。然后对各月度的日志数据进行分析,挖掘其分布的规律,进而确定目标用户在一个或多个维度下的特征标签。

[0035] 在一示例性实施例中,可以统计出日志数据中各数据的出现频次与时效性,例如可以统计各数据在全部数据中所占的比例,作为出现频次,可以统计各数据出现的时间距离当前的时间长短,作为时效性,然后结合两个指标,通过诸如相加、相乘、求平均值等计算方式得到各数据的权重,再根据各数据的权重对其进行加权计算,并将加权计算的结果确定为特征标签,例如如果日志数据为目标用户的消费水平,则可以通过对各时间分布下的消费水平进行加权计算,得到的计算结果可以作为目标用户的特征标签等等。当日志数据为文本信息时,也可以对其进行数值化处理后,进行加权计算。

[0036] 步骤S130,根据特征标签生成目标用户的画像。

[0037] 在本示例性实施例中,可以根据不同的分类标准将确定的特征标签进行聚类,以生成目标用户的画像。例如,从特征标签的内容上分类可以包括地址、购买商品类目、评价等;从特征标签的格式上分类可以包括短标签(如5个字符以内)、长标签(如5个字符以上)等等;从特征标签的形式上分类可以包括英文标签、简体中文标签、繁体中文标签等等,分类可以有其他标准,本申请对此不做特别限定。

[0038] 其中,在一示例性实施例中,步骤S120可以包括以下步骤:

[0039] 步骤S121,在各维度下,根据日志数据的时间分布确定日志数据中各数据的权重;

[0040] 步骤S122,根据日志数据中各数据的权重确定目标用户在各维度下的特征标签。

[0041] 为了有效确定目标用户的特征标签,可以在相同维度下进行日志数据的分析,例如分别对表示地址的日志数据进行分析,对支付方式的日志数据进行分析等。在其他实施例中,也可以对不同维度的日志数据进行分析,例如对地址、支付方式或者爱好等多个维度下的日志数据进行分析。

[0042] 日志数据中各数据可以是目标用户原始信息数据中具体的各个数据,例如年龄日志数据中的数据,可以是28岁、30岁等具体数据信息等,或者爱好日志数据中的数据,可以是篮球、游戏、跑步等具体数据信息等。考虑到不同时间分布的日志数据中各数据所占的权重不同,会影响目标用户特征标签的确定。因此,在本示例性实施例中,可以根据日志数据的时间分布对日志数据中各数据的权重进行设置。对日志数据中时间分布较近的数据可以

设置较大的权重，对于时间分布较远的数据可以设置较小的权重；或者对于时间分布在某一区间内的日志数据可以设置权重以某一函数类型递增，时间分布越近的权重增幅越大，而对于时间分布在较远的区间内时，日志数据的权重可以统一设置一较小的权重。权重的计算方法可以包括多种形式，可以通过特定的计算公式或函数，例如负指数函数、阶梯函数等，也可以通过人为经验设定等等。

[0043] 根据计算所得的权重，可以通过多种方式确定目标用户在各维度下的特征标签。在本示例性实施例中，日志数据中的每一数据都有其对应的权重，可以直接设置一预设标准，选择权重达到预设标准的数据作为特征标签，例如可以设置预设标准为权重最大的数据作为特征标签，或者按照权重大小进行排名后，排名前三的数据作为特征标签等。也可以确定一权重映射表，以记录数据与权重之间的映射关系。通过权重映射表，可以根据权重大小排序确定权重对应的特征标签，在对权重进行加权计算时，可以在权重映射表中查找与计算结果所对应的数据，并将其确定为特征标签等等。

[0044] 基于上述说明，在本示例性实施例中，通过获取目标用户的日志数据，并根据日志数据的时间分布确定特征标签，从而生成目标用户的画像。一方面，在本示例性实施例中，因为各数据所对应的时间分布可能不同，根据日志数据的时间分布确定特征标签，可以使生成的特征标签客观地反映出不同时间分布下用户特征数据的差异，生成较为全面的用户画像；另一方面，结合时间分布确定用户画像，可以使用户画像在生成时所考虑的因素更加丰富，且根据时间分布，能够有效确定目标用户的日志数据中各数据的变化规律，从而确定最接近当前状态的特征标签，提高用户画像生成的准确性。

[0045] 在一示例性实施例中，上述步骤S122可以包括：

[0046] 在各维度下，根据所述日志数据的各数据中权重最大的数据确定目标用户在维度下的特征标签。

[0047] 考虑到目标用户的日志数据中各数据随着时间的变化会发生改变，因此，在步骤S121中可以根据时间分布确定各数据的权重，再通过权重来确定目标用户在各维度下的特征标签，使用户画像的生成可靠性更强。在本示例性实施例中，可以将权重最大的数据确定为目标用户在各维度下的特征标签。举例说明，当目标用户的日志数据为收货地址时，如果仅考虑收货地址，随着时间的推移，目标用户可能会出现搬家或长期出差等需要更换收获地址的情况，如表1所示，用户A在2018-06之前“北京市海淀区”的收货地址的出现频率较高，而在2018-06开始出现新地址“北京市西城区”，但出现频率相比于“北京市海淀区”较低，如果仅以出现频率作为确定特征标签的标准时，用户画像的生成并不准确。因此，本示例性实施例可以根据收货地址的时间分布，为时间较近的收货地址设置较高的权重，其中权重最大的收获地址可以确定为特征标签，例如在表1中，可以确定“北京市西城区”为目标用户在收货地址维度下的特征标签。

[0048] 表1

[0049]

	收货地址	下单时间	下单次数
用户A	北京市海淀区	2017-10	11
用户A	北京市海淀区	2017-11	9
用户A	北京市海淀区	2018-02	4
用户A	北京市西城区	2018-06	7

用户A	北京市西城区	2018-08	4

[0050] 在一示例性实施例中，步骤S121可以包括以下步骤：

[0051] 步骤S210，在各维度下，按照预设的周期对日志数据进行统计，以得到所述日志数据中各数据所对应的周期序数；

[0052] 步骤S220，根据日志数据中各数据所对应的周期序数确定各数据的权重。

[0053] 其中，预设的周期可以是设定的日志数据的时间粒度，例如预设的周期可以是一周或者一个月等。根据日志数据的时间分布可以将日志数据统计至对应的周期中，例如表1中，预设的周期为月时，可以将收货地址为“北京市西城区”的日志数据统计进2018-08和2018-06月的时间周期中，将收货地址为“北京市海淀区”的日志数据分别统计进2018-02、2017-11以及2017-10月的时间周期中等，其他月的数据在此不做具体统计。周期序数是指对预设的周期进行排序得到的周期序号，可以对预设周期逆序排序，也可以正序排序，例如时间周期为月时，可以将最近的月设为最小序数，根据时间顺序，距离越远的月份，其周期的序数越大，或者也可以在某一时间区间内，从时间最远的月份开始，依次排序，直到最近的月份，例如在一年的时间区间内，时间周期为月时，可以将第一个月的序号设为1，第二个月的序号设为12等。

[0054] 在本示例性实施例中，根据预设的周期序数确定日志数据各数据的权重，可以有多种方式。其中，可以根据特定的函数计算日志数据的权重，例如负指数函数；或者可以根据经验赋值，对各周期序数计算其权重等等。举例说明，取六个月的周期序数集合，其周期序数从远到近排序可以为1～6，对各周期序数分别赋初始值$a_i=i$，$i\in[1,6]$，考虑到距离最近的月份的权重较大，因此对各周期序数的初始值求倒数，$A'=\frac{1}{a_i},\ i\in[1,6]$，然后通过以下公式，计算各周期序数的权重。

[0055] $$W=\frac{\frac{1}{a_i}}{\sum_{i=1}^{6}\frac{1}{a_i}};\qquad(1)$$

[0056] 在一示例性实施例中，日志数据的权重还可以通过以下方式确定：

[0057] 考虑到用户日志数据中各数据的时间分布对用户画像的影响，因此在计算各数据的权重时，增加时间因素。例如用户由于搬家，更换了收货地址，因此，最近的收货地址可能是最近更新的，以后经常用到。为了避免以前的历史日志数据占有较大的权重而导致特征标签提取出现问题，可以设置时间越近的数据具有较高的权重，设置时间的周期序数为x，距离当前越近的周期序数越小，越远的周期序数越大，分析两年的用户日志数据，可以使日志数据中各数据的权重满足公式：

[0058] $B(x)>B(x+1)$； (2)

[0059] 即对于距离现在越久的数据，其权重越小，而越近的数据，其权重越大。

[0060] 同时，考虑到用户偶尔出现的数据对权重计算结果的影响，例如用户偶然帮助他人购买商品使得收货地址发生变化。在计算权重时，应该将这种情况排除，可以对最近的时间周期的权重进行限制，使其低于在此之前连续出现的数据的权重之和，可以使数据的权重满足公式：

[0061] $$B(x)>\sum_{i}^{t}B(i);\qquad(3)$$

[0062] 即使日志数据中各数据满足最近数据的时间周期的权重小于近t个时间周期的权重之和，t取决于日志数据的具体特性，例如日志数据为收货地址时，设置时间过长，历史收货地址没有参考意义，时间过短，可能会使判断的特征标签出现错误，因此，综合可以设置一适中的容忍时间，如6个月。

[0063] 进一步的，如果日志数据中最近连续的时间周期里出现相同的数据，则该数据的权重可以大于之前连续T个时间周期的指标的权重。因此，对任意连续出现在两个时间周期内的数据，其权重大于之前时间周期的权重之和，考虑到时间因素对权重计算的影响程度不能过小，所以，权重的计算函数的递减速度不能过慢，即可以使数据的权重满足公式：

[0064] $$B(x) + B(x+1) > \sum_i^T B(i); \quad (4)$$

[0065] 在本示例性实施例中，还可以设置超过一定时间范围的较远的时间周期的数据，其权值可以基本不变，例如距今20个月的收货地址的权值与距今21个月的权值可以认为基本一致。因此，上述公式(4)可以不用对所有的时间周期的序数成立，而只需要对一定时间范围内的时间周期的序数满足，对于大于一定时间范围的时间周期的序数，可以满足递减函数

[0066] 根据上述说明，在一示例性实施例中，步骤S320可以包括：

[0067] 基于日志数据中各数据所对应的周期序数，通过指数函数确定各数据的权重；其中，周期序数为指数函数的指数，指数函数的底数为常数。

[0068] 在本示例性实施例中，可以通过公式(5)计算日志数据中各数据的权重：

[0069] $B(x) = a^{-bx+c}$； (5)

[0070] 即可以满足上述公式(2)、(3)、(4)所述的情况。上述x为周期序数，a、b、c为常数参数，其中，周期序数可以进行自定义设置，如果设置每个月为一个时间周期，则距今半年内，时间由远至近排序，x的取值范围可以是[1,6]，需要说明的是，该取值范围仅为示意性说明，可以根据实际需要计算的时间段，确定具体的周期序数的取值，本公开对此不做具体限定。

[0071] 在一示例性实施例中，用户画像生成方法还可以包括以下步骤：

[0072] 确定日志数据中各数据的出现频次；

[0073] 步骤S220可以包括：

[0074] 对于日志数据中任一数据D_i，通过以下公式(6)计算数据D_i的权重：

[0075] $$B(D_i) = B(S_{i1}) + B(S_{i2}) + \cdots + B(S_{im}) = freq(D_i) \cdot [e^{-S_{i1}{}^k+1} + e^{-S_{i2}{}^k+1} + \cdots + e^{-S_{im}{}^k+1}]; \quad (6)$$

[0076] 其中，B表示权重，S_{i1}、S_{i2}、…、S_{im}为数据D_i对应的周期序数，$freq(D_i)$为数据D_i在日志数据中的出现频次，k为指数常数，本示例性实施例中，指数常数k可以由上述公式(3)以及公式(4)建立的不等式进行确定，由于公式(3)要求函数的下降速率不能太大，即函数一阶导数的绝对值不能过大；而公式(4)要求函数的下降速率不能过小，即函数一阶导数的绝对值不能过小，因此，在不等式约束条件下，k值的选取由t、T的值决定。例如当t＝6，T＝22时，k＝0.66。需要说明的是，通过不等式的约束条件可以确定一k值的取值范围，基于该取值范围确定最终k值大小，例如可以将取值范围内的最小值确定为k值。

[0077] 为了使用户画像的生成更加准确，可以基于日志数据中各数据的出现频次以及时

间共同确定数据权重。其中，出现频次可以是指日志数据中各数据数量分布的指标，本示例性实施例中，各数据的出现频次可以通过多种统计方式获取，举例说明，表2示例性为某用户在2018年4月至8月的日志数据列表，示出了当日志数据为收货地址时，各数据的出现情况，其中，数据“北京市朝阳区”的出现频次的统计方法可以是“北京市朝阳区”出现的次数与5个月内所有出现的收获地址的数据的比值，如表2所示为4/10；也可以以月为时间周期，通过统计“北京市朝阳区”所在的时间周期占全部时间周期的比例，如表2所示，在时间周期8月、7月、5月、4月中均出现了“北京市朝阳区”，因此，“北京市朝阳区”的出现频次可以为4/5。根据日志数据在各时间周期内的具体情况，还可以有其他统计方式，本申请对此不做具体限制。

[0078] 表2

[0079]

日期	收货地址
2018年8月20日	北京市海淀区
2018年8月12日	北京市西城区
2018年8月5日	北京市朝阳区
2018年7月22日	北京市海淀区
2018年7月15日	北京市朝阳区
2018年6月16日	北京市海淀区
2018年5月28日	北京市西城区
2018年5月22日	北京市朝阳区
2018年4月12日	北京市西城区
2018年4月2日	北京市朝阳区

[0080] 在公式(6)，S_{i1}、S_{i2}、…、S_{im}表示数据D_i对应的周期序数，在本示例性实施例中，可以设置一时间区间，根据时间的远近程度，距当前时间最近的周期序数最小，而时间越久，周期序数越大。从而可以实现日志数据中时间分布较近的数据权重大，而时间分布较远的数据权重小，此外，考虑到数据出现频次因素的影响，如果某一数据时间分布距当前时间不是很近，但出现频次很高，其权重也可能较高，因此，公式(6)中，将各数据在日志数据中的出现频次$freq(D_i)$作为系数，来调整各数据权重的最终计算结果。

[0081] 在一示例性实施例中，步骤S110可以包括：获取目标用户在一个或多个维度下且在预设时间范围内的日志数据。

[0082] 在本示例性实施例中，在获取目标用户的日志数据时，可以对日志数据进行数据筛选。预设时间可以是对日志数据进行筛选的时间范围，例如预设时间设置为12个月，则可以在大量日志数据中获取近12个月的日志数据并进行聚合，进行用户画像的生成，预设时间可以对日志数据的获取起到过滤的作用。其中，预设时间可以根据不同维度进行设定，例如在地址维度下，考虑到用户变更地址的频率较低，可以将预设时间设置的较长(如24个月)；或者在兴趣爱好维度下，考虑到用户受网络家庭等方面影响，变更频率较高，因此，可以将预设时间设置的较短(如6个月)

[0083] 图3示意性示出本示例性实施例中另一种用户画像生成方法的流程图，首先进行步骤S310获取目标用户的日志数据，再进行步骤S320获取目标用户在一个或多个维度下的日志数据，然后可以对日志数据进行数据筛选步骤S330，确定预设时间范围内的日志数据，

再通过步骤S340根据日志数据的时间分布，确定目标用户在一个或多个维度下日志数据中各数据的权重，最后进行步骤S350根据各数据的权重确定的特征标签，从而完成用户画像的生成。

[0084] 本申请的示例性实施例还提供了一种用户画像生成装置。参照图4，该装置400可以包括，数据获取模块410，标签确定模块420以及画像生成模块430。其中，数据获取模块410用于获取目标用户在一个或多个维度下的日志数据；标签确定模块420用于根据日志数据的时间分布，确定目标用户在一个或多个维度下的特征标签；画像生成模块430用于根据特征标签生成目标用户的画像；其中，标签确定模块420可以包括：权重确定单元421，用于在各维度下，根据日志数据的时间分布确定日志数据中各数据的权重；标签处理单元422，用于根据日志数据中各数据的权重确定目标用户在各维度下的特征标签。

[0085] 在本示例性实施例中，标签确定模块可以用于在各维度下，根据日志数据的各数据中权重最大的数据确定目标用户在维度下的特征标签。

[0086] 在本示例性实施例中，权重确定单元可以包括：周期统计子单元，用于在各维度下，按照预设的周期对日志数据进行统计，以得到日志数据中各数据所对应的周期序数；权重确定子单元，用于根据日志数据中各数据所对应的周期序数确定各数据的权重。

[0087] 在本示例性实施例中，权重确定单元可以用于基于日志数据中各数据所对应的周期序数，通过指数函数确定各数据的权重；其中，周期序数为指数函数的指数，指数函数的底数为常数。

[0088] 在本示例性实施例中，用户画像生成装置还可以包括：频次确定单元，用于确定日志数据中各数据的出现频次；权重确定单元可以用于对于日志数据中任一数据D_i，通过以下公式计算数据D_i的权重：

$$B(D_i) = B(S_{i1}) + B(S_{i2}) + \cdots + B(S_{im}) = freq(D_i) \cdot [e^{-S_{i1}{}^k+1} + e^{-S_{i2}{}^k+1} + \cdots + e^{-S_{im}{}^k+1}]$$

；其中，B表示权重，S_{i1}、S_{i2}、…、S_{im}为数据D_i对应的周期序数，$freq(D_i)$为数据D_i在日志数据中的出现频次，k为指数常数。

[0089] 在本示例性实施例中，数据获取模块可以用于获取目标用户在一个或多个维度下且在预设时间范围内的日志数据。

[0090] 上述各模块/单元的具体细节已经在对应的方法部分实施例中进行了详细的描述，因此此处不再赘述。

[0091] 本申请的示例性实施例还提供了一种能够实现上述方法的电子设备。

[0092] 所属技术领域的技术人员能够理解，本申请的各个方面可以实现为系统、方法或程序产品。因此，本申请的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等)，或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。

[0093] 下面参照图5来描述根据本申请的这种示例性实施例的电子设备500。图5显示的电子设备500仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

[0094] 如图5所示，电子设备500以通用计算设备的形式表现。电子设备500的组件可以包括但不限于：上述至少一个处理单元510、上述至少一个存储单元520、连接不同系统组件(包括存储单元520和处理单元510)的总线530、显示单元540。

[0095] 其中，存储单元存储有程序代码，程序代码可以被处理单元510执行，使得处理单元510执行本说明书上述“示例性方法”部分中描述的根据本申请各种示例性实施方式的步骤。例如，处理单元510可以执行图1所示的步骤S110～S130，也可以执行图2所示的步骤S210～S220等。

[0096] 存储单元520可以包括易失性存储单元形式的可读介质，例如随机存取存储单元(RAM)521和/或高速缓存存储单元522，还可以进一步包括只读存储单元(ROM)523。

[0097] 存储单元520还可以包括具有一组(至少一个)程序模块525的程序/实用工具524，这样的程序模块525包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0098] 总线530可以为表示几类总线结构中的一种或多种，包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0099] 电子设备500也可以与一个或多个外部设备700(例如键盘、指向设备、蓝牙设备等)通信，还可与一个或者多个使得用户能与该电子设备500交互的设备通信，和/或与使得该电子设备500能与一个或多个其它计算设备进行通信的任何设备(例如路由器、调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口550进行。并且，电子设备500还可以通过网络适配器560与一个或者多个网络(例如局域网(LAN)，广域网(WAN)和/或公共网络，例如因特网)通信。如图所示，网络适配器560通过总线530与电子设备500的其它模块通信。应当明白，尽管图中未示出，可以结合电子设备500使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0100] 通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本申请实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM，U盘，移动硬盘等)中或网络上，包括若干指令以使得一台计算设备(可以是个人计算机、服务器、终端装置、或者网络设备等)执行根据本申请示例性实施例的方法。

[0101] 本申请的示例性实施例还提供了一种计算机可读存储介质，其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中，本申请的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当程序产品在终端设备上运行时，程序代码用于使终端设备执行本说明书上述“示例性方法”部分中描述的根据本申请各种示例性实施方式的步骤。

[0102] 参考图6所示，描述了根据本申请的示例性实施例的用于实现上述方法的程序产品600，其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码，并可以在终端设备，例如个人电脑上运行。然而，本申请的程序产品不限于此，在本文件中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0103] 程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导

体的系统、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0104] 计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0105] 可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF等等，或者上述的任意合适的组合。

[0106] 可以以一种或多种程序设计语言的任意组合来编写用于执行本申请操作的程序代码，程序设计语言包括面向对象的程序设计语言—诸如Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网(LAN)或广域网(WAN)，连接到用户计算设备，或者，可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0107] 此外，上述附图仅是根据本申请示例性实施例的方法所包括的处理的示意性说明，而不是限制目的。易于理解，上述附图所示的处理并不表明或限制这些处理的时间顺序。另外，也易于理解，这些处理可以是例如在多个模块中同步或异步执行的。

[0108] 应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本申请的示例性实施例，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0109] 本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本申请的其他实施例。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本申请的真正范围和精神由权利要求指出。

[0110] 应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限。

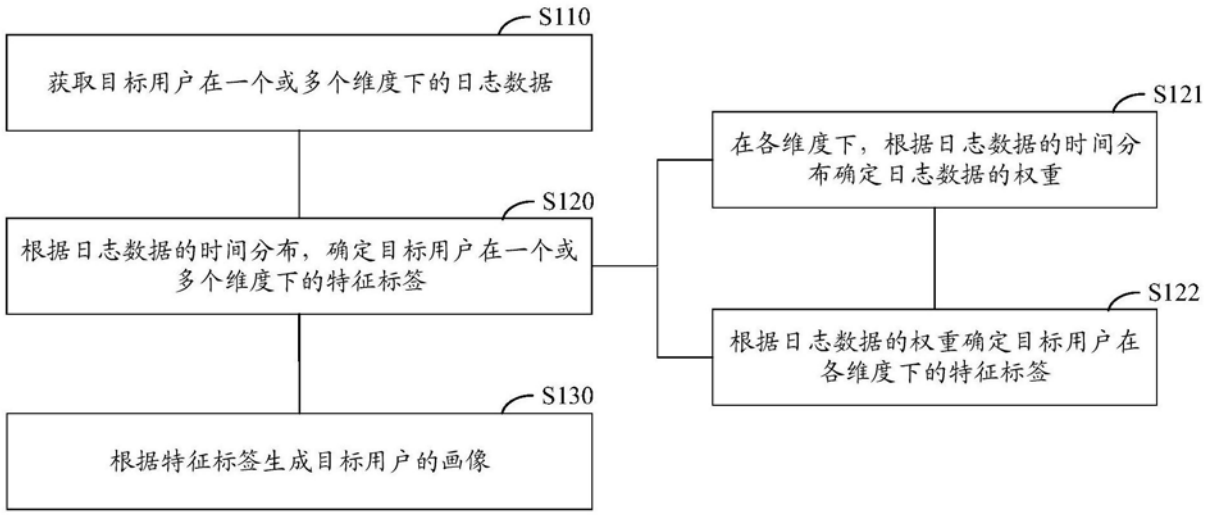

S110
获取目标用户在一个或多个维度下的日志数据
S120
根据日志数据的时间分布，确定目标用户在一个或多个维度下的特征标签
S130
根据特征标签生成目标用户的画像
S121
在各维度下，根据日志数据的时间分布确定日志数据的权重
S122
根据日志数据的权重确定目标用户在各维度下的特征标签

图1

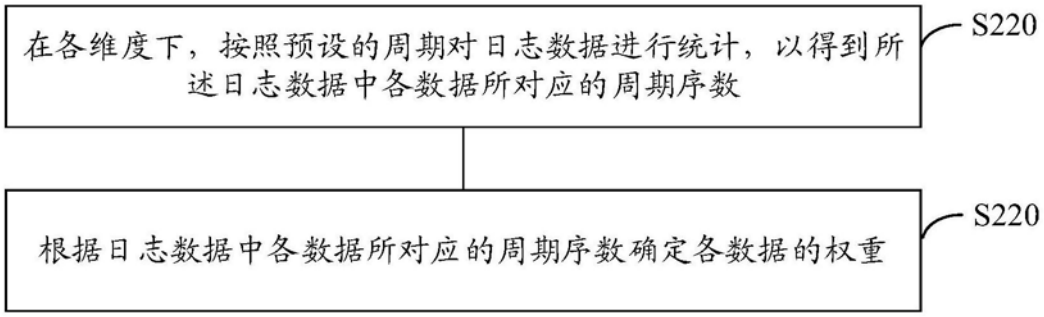

在各维度下，按照预设的周期对日志数据进行统计，以得到所述日志数据中各数据所对应的周期序数
S220
根据日志数据中各数据所对应的周期序数确定各数据的权重
S220

图2

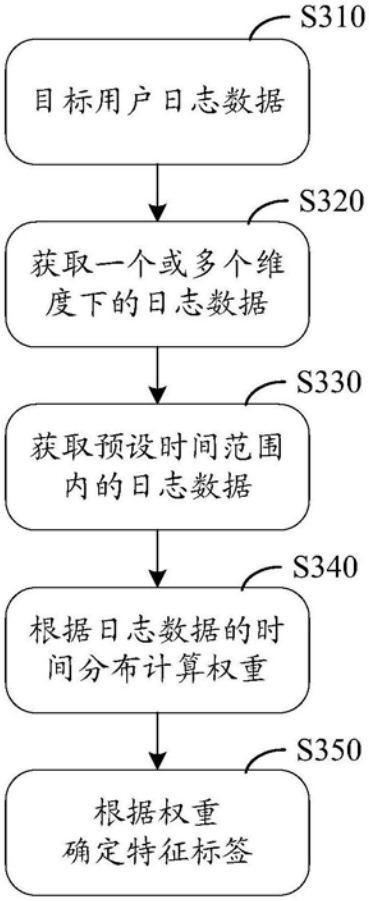

S310
目标用户日志数据
S320
获取一个或多个维度下的日志数据
S330
获取预设时间范围内的日志数据
S340
根据日志数据的时间分布计算权重
S350
根据权重确定特征标签

图3

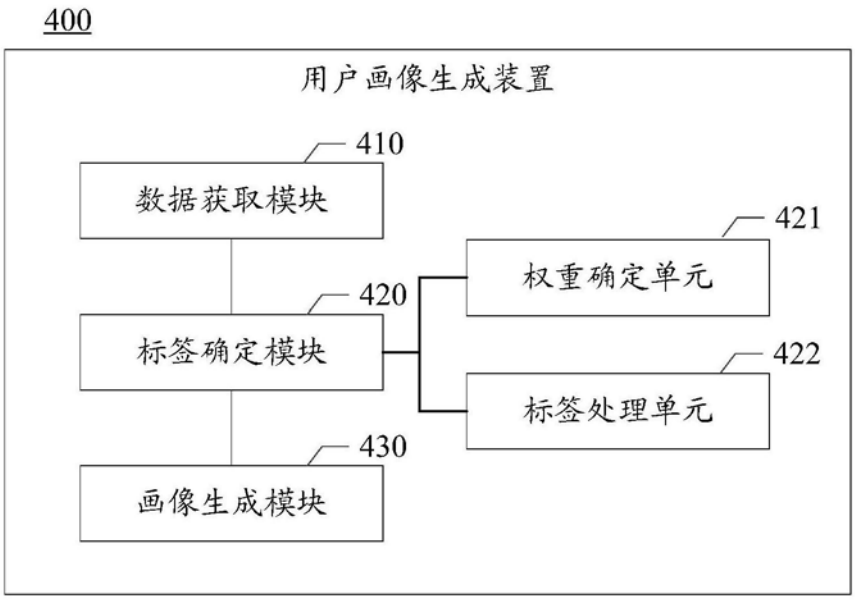

400
用户画像生成装置
410
数据获取模块
420
标签确定模块
421
权重确定单元
422
标签处理单元
430
画像生成模块

图4

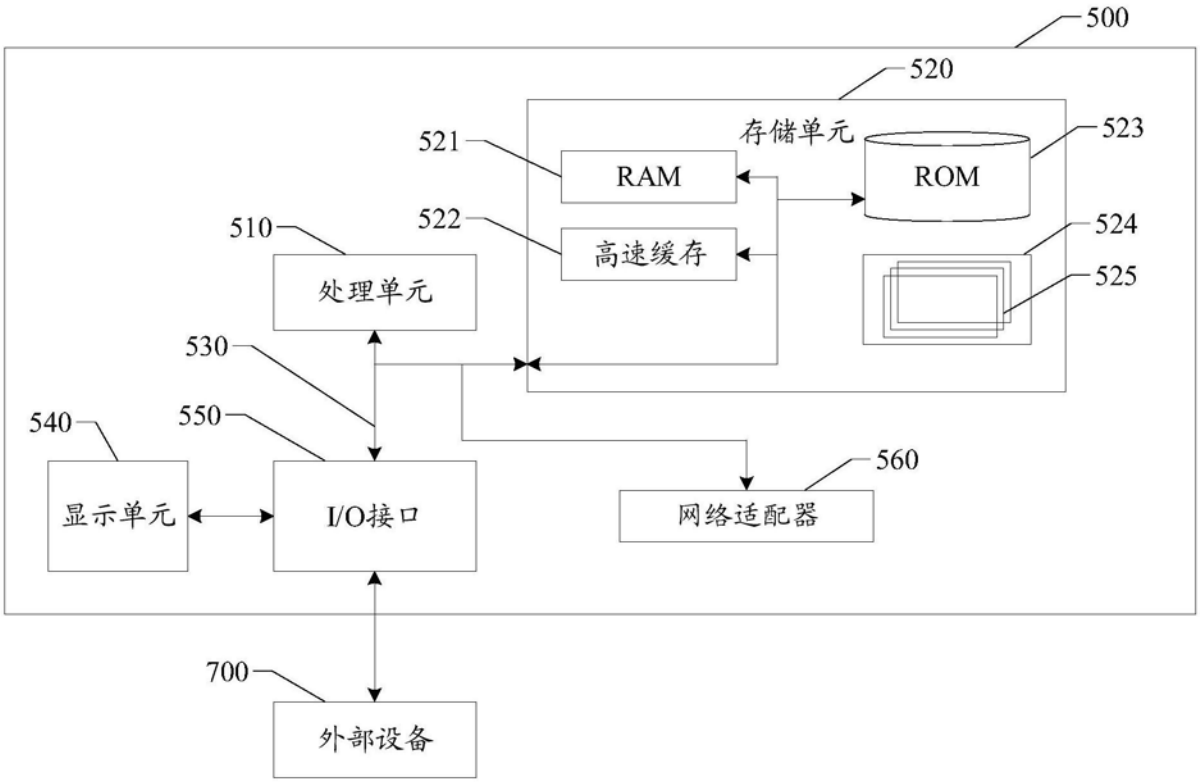

500
520
存储单元
521
RAM
523
ROM
522
高速缓存
524
525
510
处理单元
530
540
550
560
显示单元
I/O接口
网络适配器
700
外部设备

图5

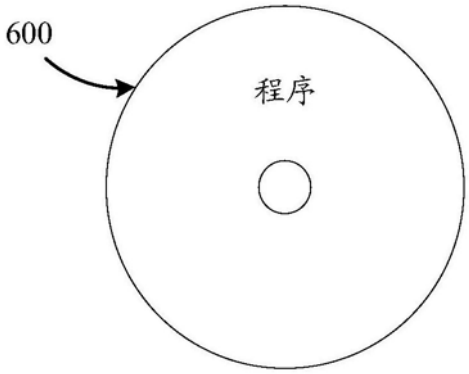

600
程序

图6