



(12) 发明专利

(10) 授权公告号 CN 111126649 B

(45) 授权公告日 2023.08.11

(21) 申请号 201811290050.2

G06F 18/214 (2023.01)

(22) 申请日 2018.10.31

(56) 对比文件

(65) 同一申请的已公布的文献号  
申请公布号 CN 111126649 A

US 2011231241 A1, 2011.09.22

CN 106021341 A, 2016.10.12

CN 108446382 A, 2018.08.24

(43) 申请公布日 2020.05.08

CN 107222526 A, 2017.09.29

(73) 专利权人 北京字节跳动网络技术有限公司  
地址 100041 北京市石景山区实兴大街30  
号院3号楼2层B-0035房间

CN 108109016 A, 2018.06.01

CN 104574124 A, 2015.04.29

US 2018005271 A1, 2018.01.04

(72) 发明人 谷长胜 张利华

CN 107526810 A, 2017.12.29

CN 105631711 A, 2016.06.01

(74) 专利代理机构 上海光栅知识产权代理有限公司 31340  
专利代理师 关浩 马雯雯

CN 103235893 A, 2013.08.07

CN 106033583 A, 2016.10.19

CN 108596755 A, 2018.09.28

(51) Int. Cl.

JP 2014095949 A, 2014.05.22

CN 105208113 A, 2015.12.30

G06F 16/9535 (2019.01)

G06N 20/00 (2019.01)

G06F 18/27 (2023.01)

审查员 方萍

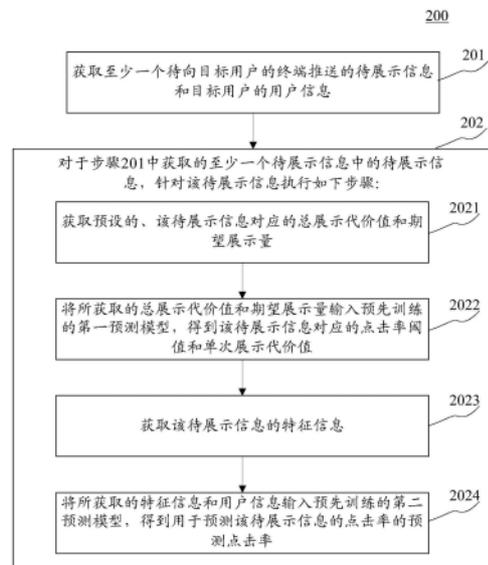
权利要求书2页 说明书12页 附图5页

(54) 发明名称

用于生成信息的方法和装置

(57) 摘要

本申请实施例公开了用于生成信息的方法和装置。该方法的一具体实施方式包括：获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息；对于所获取的至少一个待展示信息中的待展示信息，获取预设的、该待展示信息对应的总展示代价值和期望展示量；将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型，得到该待展示信息对应的点击率阈值和单次展示代价值；获取该待展示信息的特征信息；将所获取的特征信息和用户信息输入预先训练的第二预测模型，得到用于预测该待展示信息的点击率的预测点击率。该实施方式可以提高信息生成的准确性，有助于提高信息推送的针对性。



1. 一种用于生成信息的方法,包括:

获取至少一个待向目标用户的终端推送的待展示信息和所述目标用户的用户信息;

对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;

获取该待展示信息的特征信息;将所获取的特征信息和所述用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率;

从所述至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息;所述预设条件包括以下至少一种:单次展示代价值是所得到的单次展示代价值中的最大值、单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值;

将所提取的待展示信息推送至所述终端。

2. 根据权利要求1所述的方法,其中,所述第一预测模型通过如下步骤训练得到:

获取第一训练样本集合,其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值;

利用机器学习方法,将所述第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、标注单次展示代价值作为期望输出,训练得到第一预测模型。

3. 根据权利要求1或2所述的方法,其中,所述第一预测模型是基于线性回归模型训练得到的模型。

4. 根据权利要求1或2所述的方法,其中,所述第二预测模型通过如下步骤训练得到:

获取第二训练样本集合,其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息;

利用机器学习方法,将所述第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。

5. 一种用于生成信息的装置,包括:

获取单元,被配置成获取至少一个待向目标用户的终端推送的待展示信息和所述目标用户的用户信息;

生成单元,被配置成对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;

获取该待展示信息的特征信息;将所获取的特征信息和所述用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率;

提取单元,被配置成从所述至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息;所述预设条件包括以下至少一种:单次展示代价值是所得到的单次展示代价值中的最大值、单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值;

推送单元,被配置成将所提取的待展示信息推送至所述终端。

6. 根据权利要求5所述的装置,其中,所述第一预测模型通过如下步骤训练得到:

获取第一训练样本集合,其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值;

利用机器学习方法,将所述第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、标注单次展示代价值作为期望输出,训练得到第一预测模型。

7. 根据权利要求5或6所述的装置,其中,所述第一预测模型是基于线性回归模型训练得到的模型。

8. 根据权利要求5或6所述的装置,其中,所述第二预测模型通过如下步骤训练得到:

获取第二训练样本集合,其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息;

利用机器学习方法,将所述第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。

9. 一种服务器,包括:

一个或多个处理器;

存储装置,其上存储有一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-4中任一所述的方法。

10. 一种计算机可读介质,其上存储有计算机程序,其中,该程序被处理器执行时实现如权利要求1-4中任一所述的方法。

## 用于生成信息的方法和装置

### 技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及用于生成信息的方法和装置。

### 背景技术

[0002] 随着互联网技术的发展,用户通过终端浏览信息越来越频繁,现有技术通常针对某一信息,计算出某一时间段内该信息的期望展示量和该时间段内展示的信息的总展示量的比值,将计算出的比值作为推送信息的频率,根据频率向用户推送信息,从而该信息的实际展示量达到期望展示量。

### 发明内容

[0003] 本申请实施例提出了用于生成信息的方法和装置。

[0004] 第一方面,本申请实施例提供了一种用于生成信息的方法,该方法包括:获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息;对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0005] 在一些实施例中,该方法还包括:从至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息;将所提取的待展示信息推送至终端。

[0006] 在一些实施例中,预设条件包括以下至少一种:单次展示代价值是所得到的单次展示代价值中的最大值;单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值。

[0007] 在一些实施例中,第一预测模型通过如下步骤训练得到:获取第一训练样本集合,其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值;利用机器学习方法,将第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、标注单次展示代价值作为期望输出,训练得到第一预测模型。

[0008] 在一些实施例中,第一预测模型是基于线性回归模型训练得到的模型。

[0009] 在一些实施例中,第二预测模型通过如下步骤训练得到:获取第二训练样本集合,其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息;利用机器学习方法,将第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。

[0010] 第二方面,本申请实施例提供了一种用于生成信息的装置,该装置包括:获取单元,被配置成获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息;生成单元,被配置成对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0011] 在一些实施例中,该装置还包括:提取单元,被配置成从至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息;推送单元,被配置成将所提取的待展示信息推送至终端。

[0012] 在一些实施例中,预设条件包括以下至少一种:单次展示代价值是所得到的单次展示代价值中的最大值;单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值。

[0013] 在一些实施例中,第一预测模型通过如下步骤训练得到:获取第一训练样本集合,其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值;利用机器学习方法,将第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、标注单次展示代价值作为期望输出,训练得到第一预测模型。

[0014] 在一些实施例中,第一预测模型是基于线性回归模型训练得到的模型。

[0015] 在一些实施例中,第二预测模型通过如下步骤训练得到:获取第二训练样本集合,其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息;利用机器学习方法,将第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。

[0016] 第三方面,本申请实施例提供了一种服务器,该服务器包括:一个或多个处理器;存储装置,其上存储有一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

[0017] 第四方面,本申请实施例提供了一种计算机可读介质,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如第一方面中任一实现方式描述的方法。

[0018] 本申请实施例提供的用于生成信息的方法和装置,通过获取至少一个待展示信息和目标用户的用户信息,然后获取待展示信息对应的总展示代价值和期望展示量,使用第一预测模型和第二预测模型,得到待展示信息对应的点击率阈值、单次展示代价值和预测点击率,从而可以提高信息生成的准确性,有助于利用模型输出的数据提高信息推送的针对性。

## 附图说明

[0019] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

- [0020] 图1是本申请的一个实施例可以应用于其中的示例性系统架构图；
- [0021] 图2是根据本申请实施例的用于生成信息的方法的一个实施例的流程图；
- [0022] 图3是根据本申请实施例的用于生成信息的方法的一个应用场景的示意图；
- [0023] 图4是根据本申请实施例的用于生成信息的方法的又一个实施例的流程图；
- [0024] 图5是根据本申请实施例的用于生成信息的装置的一个实施例的结构示意图；
- [0025] 图6是适于用来实现本申请实施例的服务器的计算机系统的结构示意图。

### 具体实施方式

[0026] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

[0027] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0028] 图1示出了可以应用本申请实施例的用于生成信息的方法或用于生成信息的装置的示例性系统架构100。

[0029] 如图1所示，系统架构100可以包括终端设备101、102、103，网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0030] 用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用，例如网页浏览器应用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0031] 终端设备101、102、103可以是硬件，也可以是软件。当终端设备101、102、103为硬件时，可以是具有显示屏的各种电子设备，包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。当终端设备101、102、103为软件时，可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

[0032] 服务器105可以是提供各种服务的服务器，例如对终端设备101、102、103上展示的信息提供支持的后台信息处理服务器。后台信息处理服务器可以对获取的待展示信息和用户信息等信息进行处理，并生成处理结果（例如点击率阈值、单次展示代价值和预测点击率）。

[0033] 需要说明的是，本申请实施例所提供的用于生成信息的方法一般由服务器105执行，相应地，用于生成信息的装置一般设置于服务器105中。

[0034] 需要说明的是，服务器可以是硬件，也可以是软件。当服务器为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

[0035] 应该理解，图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器。

[0036] 继续参考图2，示出了根据本申请的用于生成信息的方法的一个实施例的流程

200。该用于生成信息的方法,包括以下步骤:

[0037] 步骤201,获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息。

[0038] 在本实施例中,用于生成信息的方法的执行主体(例如图1所示的服务器)可以通过有线连接方式或者无线连接方式从远程或从本地获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息。其中,待展示信息可以包括但不限于以下至少一项:图片、文字、音频、视频、链接地址等。目标用户可以是待利用其使用的终端(例如图1所示的终端设备)浏览待展示信息的用户。目标用户的用户信息可以用于表征目标用户的特征,目标用户的特征包括但不限于以下至少一种:目标用户的性别、年龄、兴趣等。

[0039] 步骤202,对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0040] 在本实施例中,对于步骤201中获取的至少一个待展示信息中的待展示信息,上述执行主体可以针对该待展示信息执行如下步骤:

[0041] 步骤2021,获取预设的、该待展示信息对应的总展示代价值和期望展示量。

[0042] 其中,总展示代价值用于表征待展示信息的提供者为了向用户展示待展示信息所付出的总的代价(例如价格、积分值等)。期望展示量为预设的、期望浏览待展示信息的用户的数量,或者是上述执行主体推送待展示信息的次数。

[0043] 步骤2022,将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值。

[0044] 其中,单次展示代价值用于表征在用户使用的终端上展示一次待展示信息,待展示信息的提供者(例如向用户提供的待展示信息指示的物品的所有者)所付出的代价。第一预测模型用于表征总展示代价值、期望展示量与点击率阈值、单次展示代价值的对应关系。具体地,作为示例,第一预测模型可以是技术人员预先基于对大量的总展示代价值、期望展示量与点击率阈值、单次展示代价值的统计而预先制定的、存储有多个总展示代价值、期望展示量与对应的点击率阈值、单次展示代价值的对应关系表;也可以为基于预设的训练样本,利用机器学习方法对初始模型(例如广义线性回归模型、局部加权线性回归模型、神经网络等)进行训练后得到的模型。通过使用第一预测模型,可以根据不同的待展示信息的总展示代价值和期望展示量,得到不同的点击率阈值和单次展示代价值,有助于为确定是否向目标用户推送待展示信息提供参考。

[0045] 在本实施例的一些可选的实现方式中,上述执行主体或其他电子设备可以按照如下步骤训练得到第一预测模型:

[0046] 首先,获取第一训练样本集合。其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值。

[0047] 然后,利用机器学习方法,将第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、

标注单次展示代价值作为期望输出,训练得到第一预测模型。

[0048] 具体地,上述第一预测模型可以是对初始模型进行训练得到的模型。初始模型可以是线性回归模型、神经网络模型等。初始模型可以设置有初始参数,参数在训练过程中可以被不断地调整。训练第一预测模型的执行主体可以基于预设的损失函数计算损失值,根据损失值确定初始模型是否训练完成。在这里,需要说明的是,损失值可以用于表征实际输出与期望输出之间的差异。实践中,可以采用预设的各种损失函数计算实际输出相对于标注的输出的损失值。例如,可以采用平方损失函数计算损失值。

[0049] 在本实施例的一些可选的实现方式中,第一预测模型可以是基于线性回归模型训练得到的模型。线性回归是利用数理统计中回归分析,来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法,方法简单,运用广泛。

[0050] 实践中,上述第一预测模型可以包括两个线性回归模型,即第一线性回归模型和第二线性回归模型。其中,第一线性回归模型可以用于表征总展示代价值、期望展示量与点击率阈值的对应关系。第二线性回归模型可以用于表征总展示代价值、期望展示量与单次展示代价值的对应关系。

[0051] 作为示例,假设第一线性回归模型和第二线性回归模型分别如式(1)和式(2)所示:

$$[0052] \quad f_1(S,C) = \text{CTR\_th} \quad (1)$$

$$[0053] \quad f_2(S,C) = \text{bid} \quad (2),$$

[0054] 其中,S为期望展示量,C为总展示代价值,CTR\_th为点击率阈值,bid为单次展示代价值。在训练第一预测模型时,可以将样本期望展示量、样本总展示代价值作为初始第一线性回归模型的输入,将标注点击率阈值作为初始第一线性回归模型的期望输出,训练得到第一线性回归模型。将样本期望展示量、样本总展示代价值作为初始第二线性回归模型的输入,将标注单次展示代价值作为初始第二线性回归模型的期望输出,训练得到第二线性回归模型。

[0055] 在本示例中,假设上述初始第一线性回归模型和初始第二线性回归模型如下式(3)和式(4)所示:

$$[0056] \quad \alpha_1 \times S + \alpha_2 \times C = \text{CTR\_th} \quad (3)$$

$$[0057] \quad \alpha_3 \times S + \alpha_4 \times C = \text{bid} \quad (4),$$

[0058] 其中, $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$ 、 $\alpha_4$ 分别为初始第一线性回归模型和初始第二线性回归模型的参数。对初始第一线性回归模型和初始第二线性回归模型进行训练后,可以确定参数 $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$ 、 $\alpha_4$ 的参数值,从而得到训练后的第一线性回归模型和第二线性回归模型。当上述执行主体向第一预测模型输入总展示代价值C和期望展示量S时,由训练后的第一线性回归模型和第二线性回归模型分别计算出点击率阈值CTR\_th和单次展示代价值bid。

[0059] 可选地,上述第一线性回归模型和第二线性回归模型的自变量和因变量可以互换。即第一线性回归模型和第二线性回归模型可以分别如式(5)和式(6)所示:

$$[0060] \quad f_1(\text{CTR\_th}, \text{bid}) = C \quad (5)$$

$$[0061] \quad f_2(\text{CTR\_th}, \text{bid}) = S \quad (6),$$

[0062] 在训练第一预测模型时,可以将标注点击率阈值、标注单次展示代价值作为初始第一线性回归模型的输入,将样本总展示代价值作为初始第一线性回归模型的期望输出,

训练得到第一线性回归模型。将标注点击率阈值、标注单次展示代价值作为初始第二线性回归模型的输入,将样本期望展示量作为初始第二线性回归模型的期望输出,训练得到第二线性回归模型。

[0063] 在本示例中,假设上述初始第一线性回归模型和初始第二线性回归模型如下式(7)和式(8)所示:

$$[0064] \quad \beta_1 \times \text{CTR\_th} + \beta_2 \times \text{bid} = C \quad (7)$$

$$[0065] \quad \beta_3 \times \text{CTR\_th} + \beta_4 \times \text{bid} = S \quad (8),$$

[0066] 其中, $\beta_1$ 、 $\beta_2$ 、 $\beta_3$ 、 $\beta_4$ 分别为初始第一线性回归模型和初始第二线性回归模型的参数。对初始第一线性回归模型和初始第二线性回归模型进行训练后,可以确定参数 $\beta_1$ 、 $\beta_2$ 、 $\beta_3$ 、 $\beta_4$ 的参数值,从而得到训练后的第一线性回归模型和第二线性回归模型。

[0067] 当上述执行主体向第一预测模型输入总展示代价值C和期望展示量S时,可以求解由上述式(7)和式(8)组成的方程组,从而得到点击率阈值CTR\_th和单次展示代价值bid。

[0068] 步骤2023,获取该待展示信息的特征信息。

[0069] 其中,特征信息用于表征待展示信息的特征。待展示信息的特征可以包括但不限于以下至少一种:待展示信息的标题、类型、链接地址等。

[0070] 步骤2024,将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0071] 其中,点击率又称点击通过率(CTR,Click Through Rate),即某展示信息的实际点击次数除以展示信息的展示量(例如推送展示信息的次数、接收展示信息的终端的数量等)的结果。预测点击率可以用于表征该待展示信息被目标用户点击的概率,即预测点击率越大,表示该待展示信息被目标用户点击的可能性越大。

[0072] 第二预测模型用于表征特征信息和用户信息与预测点击率的对应关系。具体地,作为示例,第二预测模型可以是技术人员预先基于对大量的待展示信息的特征信息、用户信息与预测点击率的统计而预先制定的、存储有多个特征信息、用户信息与对应的预测点击率的对应关系表;也可以为基于预设的训练样本,利用机器学习方法对初始模型(例如FM(Factorization Machine,因子分解机)模型、FFM(Field-aware Factorization Machine,场感知因子分解机)模型、神经网络模型等)进行训练后得到的模型。通过使用第二预测模型,可以准确地预测目标用户点击待展示信息的可能性,从而有助于提高向目标用户推送待展示信息的针对性。

[0073] 在本实施例的一些可选的实现方式中,上述执行主体或其他电子设备可以按照如下步骤训练得到第二预测模型:

[0074] 首先,获取第二训练样本集合。其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息。作为示例,标注信息可以是数字,例如“0”表示样本用户没有点击样本展示信息,“1”表示样本用户点击了样本展示信息。

[0075] 然后,利用机器学习方法,将第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。作为示例,第二预测模型的输出数据可以是介于0和1之间的数值,该数值越接近1,表征特征信息指示的待展示信息被用户点击的可能性越大。

[0076] 具体地,上述第二预测模型可以是对初始模型进行训练得到的模型。初始模型可以包括FM模型、神经网络模型等。初始模型可以设置有初始参数,参数在训练过程中可以被不断地调整。训练第一预测模型的执行主体可以基于预设的损失函数计算损失值,根据损失值确定初始模型是否训练完成。在这里,需要说明的是,损失值可以用于表征实际输出与期望输出之间的差异。实践中,可以采用预设的各种损失函数计算实际输出相对于标注的输出的损失值。例如,可以采用对数损失函数、交叉熵损失函数等计算损失值。

[0077] 继续参见图3,图3是根据本实施例的用于生成信息的方法的应用场景的一个示意图。在图3的应用场景中,服务器301首先从本地获取三个待向目标用户的终端推送的待展示信息3021、3022、3023,以及目标用户303的用户信息304。其中,用户信息304包括目标用户303的性别、年龄等信息。然后,服务器301获取预设的、每个待展示信息对应的总展示代价值3051(例如“1000”)、3052(例如“2000”)、3053(例如“2500”),和期望展示量3061(例如“10000”)、3062(例如“20000”)、3063(例如“10000”)。接着,服务器301将所获取的总展示代价值和期望展示量分别输入预先训练的第一预测模型307,得到每个待展示信息对应的点击率阈值3071(例如“10%”)、3072(例如“5.9%”)、3073(例如“16%”)和单次展示代价值3081(例如“1.2”)、3082(例如“0.9”)、3083(例如“1.5”)。再然后,服务器301获取每个待展示信息的特征信息(例如图中的3091、3092、3093),将所获取的特征信息和用户信息304输入预先训练的第二预测模型310,得到用于预测每个待展示信息的点击率的预测点击率3101(例如“11%”)、3102(例如“5%”)、3103(例如“13%”)。

[0078] 本申请的上述实施例提供的方法,通过使用第一预测模型和第二预测模型,得到待展示信息对应的点击率阈值、单次展示代价值和预测点击率,从而可以提高信息生成的准确性,有助于提高信息推送的针对性。

[0079] 进一步参考图4,其示出了用于生成信息的方法的又一个实施例的流程400。该用于生成信息的方法的流程400,包括以下步骤:

[0080] 步骤401,获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息。

[0081] 在本实施例中,步骤401与图2对应实施例中的步骤201基本一致,这里不再赘述。

[0082] 步骤402,对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0083] 在本实施例中,步骤402与图2对应实施例中的步骤202基本一致,这里不再赘述。

[0084] 步骤403,从至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息。

[0085] 在本实施例中,用于生成信息的方法的执行主体(例如图1所示的服务器)可以从上述至少一个待展示信息中,确定对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息。点击率阈值可以用于与预测点击率进行比较,当某待展示信息对应的预测点击率大于等于点击率阈值时,表示该待展示信息被目标用户点击的可能性较大。预设条件可以是技术人员预先设置的、用于从至少一个待展示信

息中选择待展示信息的条件。

[0086] 在本实施例的一些可选的实现方式中,预设条件可以包括但不限于以下至少一种:

[0087] 单次展示代价值是所得到的单次展示代价值中的最大值;

[0088] 单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值。

[0089] 其中,预设位置可以是单个位置,也可以是多个位置。例如,预设位置可以是将所得到的各个单次代价值由大到小排序后,排在前N(N为预设的正整数)位的单次展示代价值。

[0090] 步骤404,将所提取的待展示信息推送至终端。

[0091] 在本实施例中,上述执行主体可以将步骤403中确定的待展示信息推送至上述终端。以使在终端上显示待展示信息。

[0092] 从图4中可以看出,与图2对应的实施例相比,本实施例中的用于生成信息的方法的流程400突出了从至少一个待展示信息中提取待展示信息和推送待展示信息的步骤。由此,本实施例描述的方案可以基于第一预测模型和第二预测模型的输出结果,更加具有针对性地向目标用户的终端推送信息。

[0093] 进一步参考图5,作为对上述各图所示方法的实现,本申请提供了一种用于生成信息的装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0094] 如图5所示,本实施例的用于生成信息的装置500包括:获取单元501,被配置成获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息;生成单元502,被配置成对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0095] 在本实施例中,获取单元501可以通过有线连接方式或者无线连接方式从远程或从本地获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息。其中,待展示信息可以包括但不限于以下至少一项:图片、文字、音频、视频、链接地址等。目标用户可以是待利用其使用的终端(例如图1所示的终端设备)浏览待展示信息的用户。目标用户的用户信息可以用于表征目标用户的特征,目标用户的特征包括但不限于以下至少一种:目标用户的性别、年龄、兴趣等。

[0096] 在本实施例中,对于获取单元501获取的至少一个待展示信息中的待展示信息,上述生成单元502可以针对该待展示信息执行如下步骤:

[0097] 步骤5021,获取预设的、该待展示信息对应的总展示代价值和期望展示量。

[0098] 其中,总展示代价值用于表征待展示信息的提供者为了向用户展示待展示信息所付出的代价。期望展示量为预设的、期望浏览待展示信息的用户的数量,或者是上述执行主体推送待展示信息的次数。

[0099] 步骤5022,将所获取的总展示代价值和期望展示量输入预先训练的第一预测模

型,得到该待展示信息对应的点击率阈值和单次展示代价值。

[0100] 其中,第一预测模型用于表征总展示代价值、期望展示量与点击率阈值、单次展示代价值的对应关系。具体地,作为示例,第一预测模型可以是技术人员预先基于对大量的总展示代价值、期望展示量与点击率阈值、单次展示代价值的统计而预先制定的、存储有多个总展示代价值、期望展示量与对应的点击率阈值、单次展示代价值的对应关系表;也可以为基于预设的训练样本,利用机器学习方法对初始模型(例如广义线性回归模型、局部加权线性回归模型、神经网络等)进行训练后得到的模型。通过使用第一预测模型,可以根据不同的待展示信息的总展示代价值和期望展示量,得到不同的点击率阈值和单次展示代价值,有助于为确定是否向目标用户推送待展示信息提供参考。

[0101] 步骤5023,获取该待展示信息的特征信息。

[0102] 其中,特征信息用于表征待展示信息的特征。待展示信息的特征可以包括但不限于以下至少一种:待展示信息的标题、类型、链接地址等。

[0103] 步骤5024,将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0104] 其中,点击率又称点击通过率(CTR,Click Through Rate),即某展示信息的实际点击次数除以展示信息的展示量(例如推送展示信息的次数、接收展示信息的终端的数量等)的结果。预测点击率可以用于表征该待展示信息被目标用户点击的概率,即预测点击率越大,表示该待展示信息被目标用户点击的可能性越大。

[0105] 第二预测模型用于表征特征信息和用户信息与预测点击率的对应关系。具体地,作为示例,第二预测模型可以是技术人员预先基于对大量的待展示信息的特征信息、用户信息与预测点击率的统计而预先制定的、存储有多个特征信息、用户信息与对应的预测点击率的对应关系表;也可以为基于预设的训练样本,利用机器学习方法对初始模型(例如FM(Factorization Machine,因子分解机)模型、FFM(Field-aware Factorization Machine,场感知因子分解机)模型、神经网络模型等)进行训练后得到的模型。通过使用第二预测模型,可以准确地预测目标用户点击待展示信息的可能性,从而有助于提高向目标用户推送待展示信息的针对性。

[0106] 在本实施例的一些可选的实现方式中,该装置500还可以包括:提取单元(图中未示出),被配置成从至少一个待展示信息中,提取对应的预测点击率大于等于对应的点击率阈值且对应的单次展示代价值符合预设条件的待展示信息;推送单元(图中未示出),被配置成将所提取的待展示信息推送至终端。

[0107] 在本实施例的一些可选的实现方式中,预设条件可以包括以下至少一种:单次展示代价值是所得到的单次展示代价值中的最大值;单次展示代价值是按照所得到的单次展示代价值的大小排列后,排在预设位置处的单次展示代价值。

[0108] 在本实施例的一些可选的实现方式中,第一预测模型通过如下步骤训练得到:获取第一训练样本集合,其中,第一训练样本包括预先针对样本待展示信息设置的样本总展示代价值、样本期望展示量,以及对应的标注点击率阈值、标注单次展示代价值;利用机器学习方法,将第一训练样本集合中的样本总展示代价值和样本期望展示量作为输入,将输入的样本总展示代价值、样本期望展示量对应的标注点击率阈值、标注单次展示代价值作为期望输出,训练得到第一预测模型。

[0109] 在一些实施例中,第一预测模型是基于线性回归模型训练得到的模型。

[0110] 在一些实施例中,第二预测模型通过如下步骤训练得到:获取第二训练样本集合,其中,第二训练样本包括预先获取的样本展示信息的特征信息、浏览样本展示信息的样本用户的样本用户信息,以及对应的、用于表征样本用户是否点击样本展示信息的标注信息;利用机器学习方法,将第二训练样本集合中的特征信息和样本用户信息作为输入,将输入的特征信息和样本用户信息对应的标注信息作为期望输出,训练得到第二预测模型。

[0111] 本申请的上述实施例提供的装置,通过使用第一预测模型和第二预测模型,得到待展示信息对应的点击率阈值、单次展示代价值和预测点击率,从而可以提高信息生成的准确性,有助于提高信息推送的针对性。

[0112] 下面参考图6,其示出了适于用来实现本申请实施例的服务器的计算机系统600的结构示意图。图6示出的服务器仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0113] 如图6所示,计算机系统600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0114] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;包括诸如液晶显示器(LCD)等以及扬声器等的输出部分607;包括硬盘等的存储部分608;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器610上,以便于从其上读出的计算机程序根据需要被安装入存储部分608。

[0115] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时,执行本申请的方法中限定的上述功能。

[0116] 需要说明的是,本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读介质或者是上述两者的任意组合。计算机可读介质例如可以是——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。计算机可读介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读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。而在本申请中,计算机可读的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,其中承载了计算机可读的程序代码。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。计算机可读的信号介质还可以是计算机可读介质以外的任何计算机可读介质,该计算机可读介质可以发送、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。计算机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、电线、光缆、RF等等,或者上述的任意合适的组合。

[0117] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0118] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0119] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包括获取单元和生成单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息的单元”。

[0120] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的服务器中所包含的;也可以是单独存在,而未装配入该服务器中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该服务器执行时,使得该服务器:获取至少一个待向目标用户的终端推送的待展示信息和目标用户的用户信息;对于所获取的至少一个待展示信息中的待展示信息,获取预设的、该待展示信息对应的总展示代价值和期望展示量;将所获取的总展示代价值和期望展示量输入预先训练的第一预测模型,得到该待展示信息对应的点击率阈值和单次展示代价值;获取该待展示信息的特征信息;将所获取的特征信息和用户信息输入预先训练的第二预测模型,得到用于预测该待展示信息的点击率的预测点击率。

[0121] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功

能的技术特征进行互相替换而形成的技术方案。

100

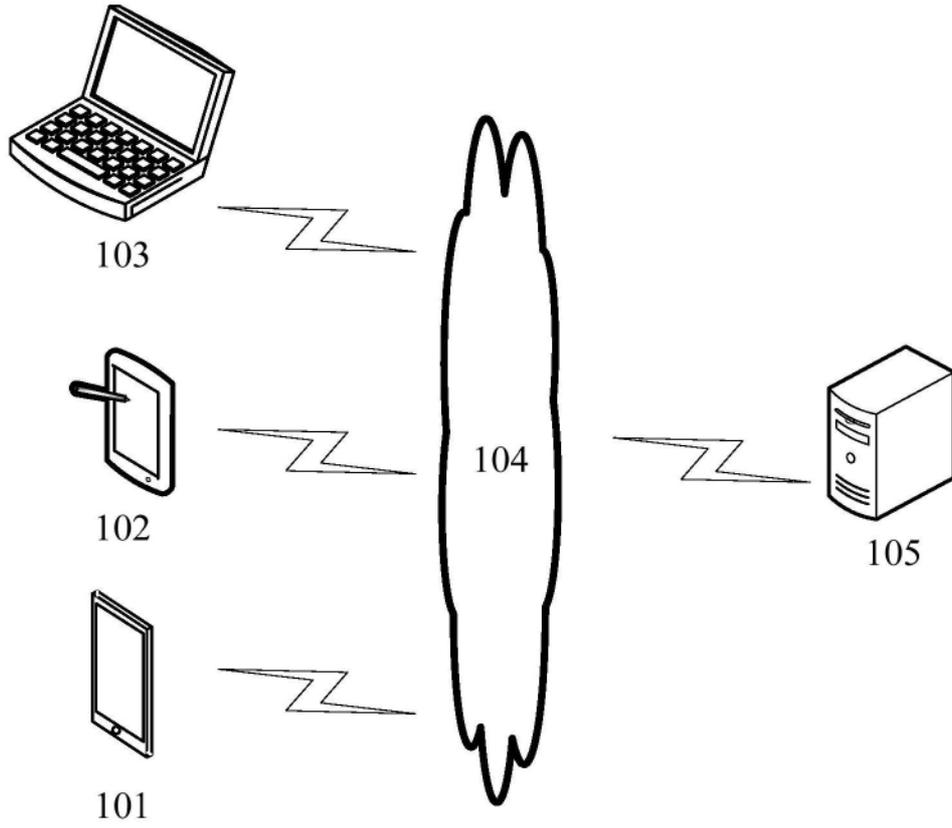


图1

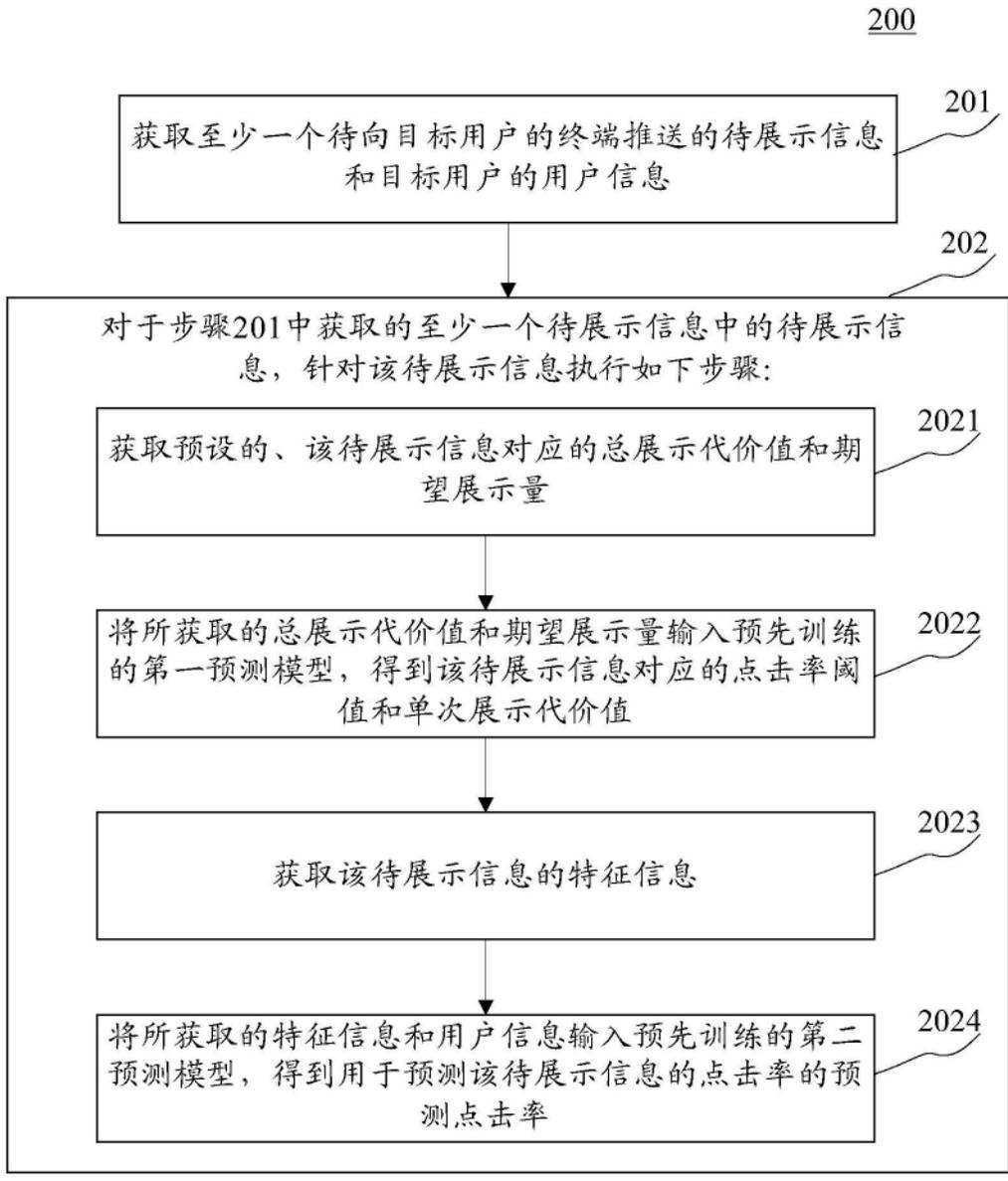


图2

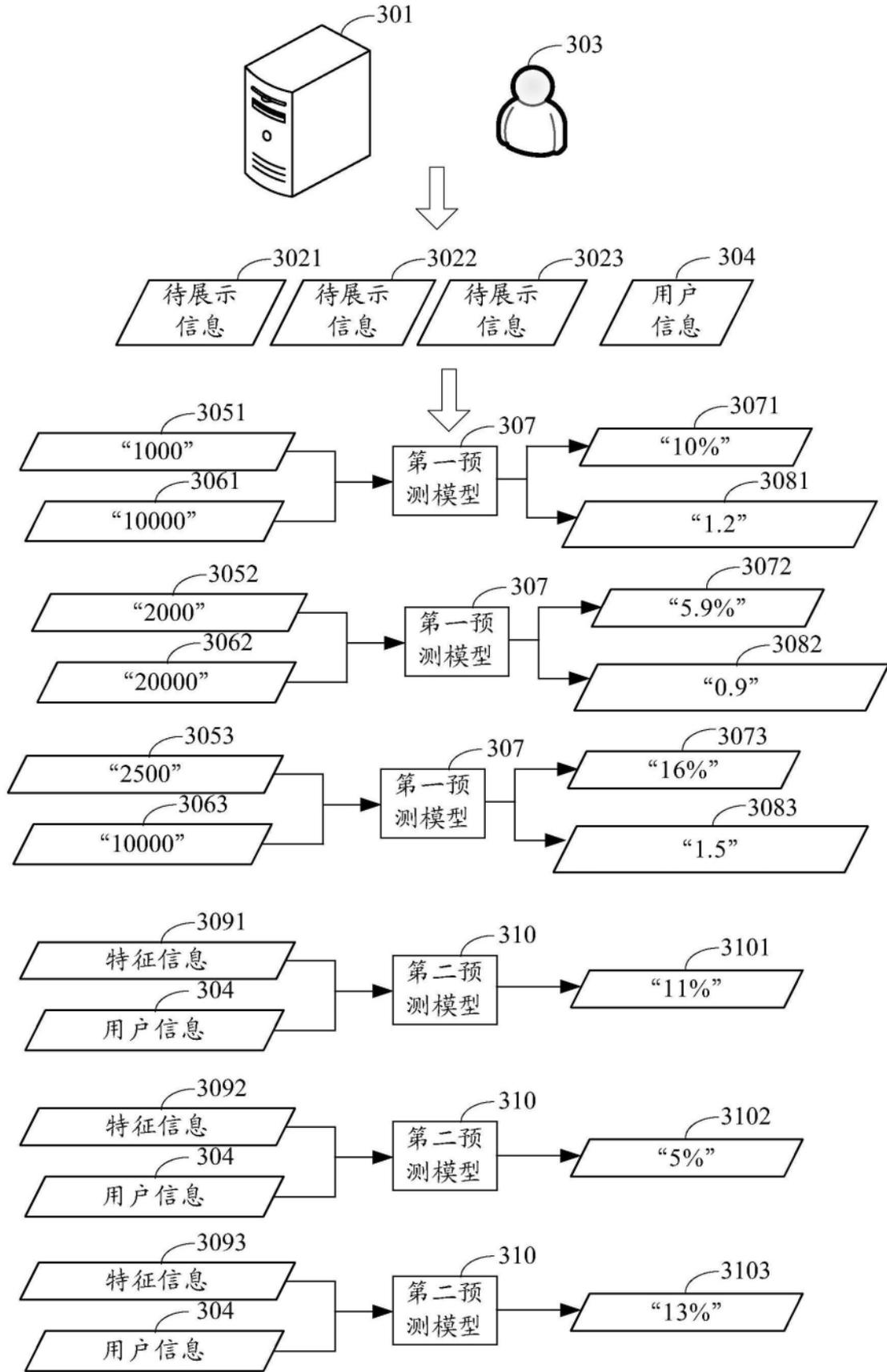


图3

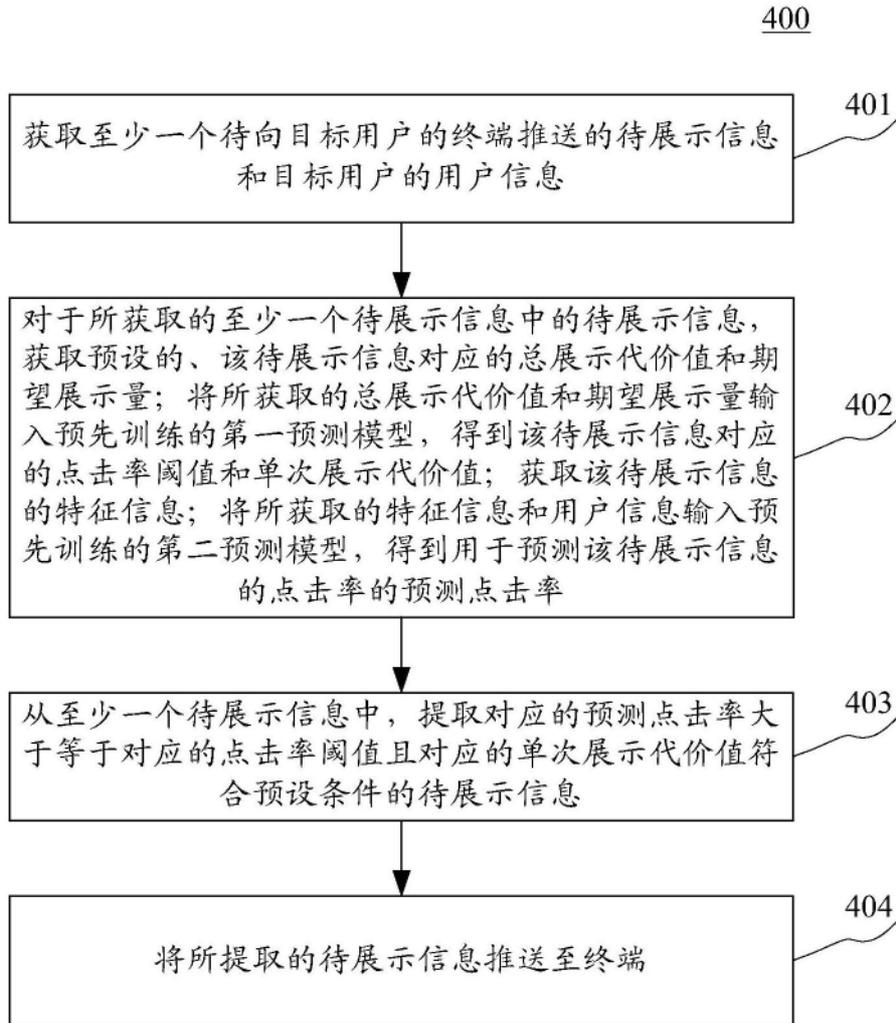


图4

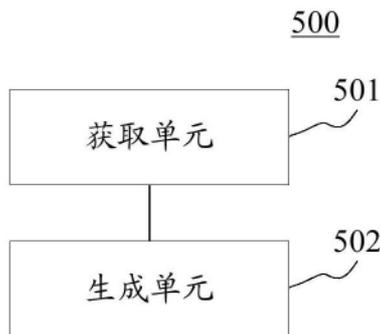


图5

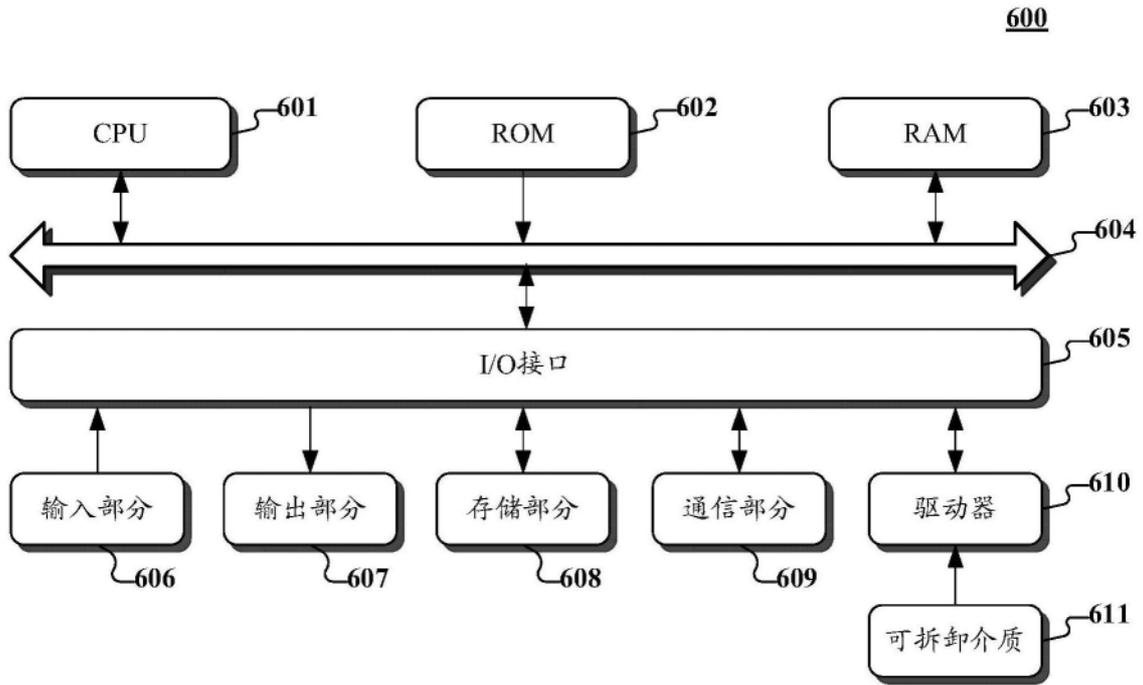


图6