



(12)发明专利申请

(10)申请公布号 CN 109241427 A

(43)申请公布日 2019.01.18

(21)申请号 201811024963.X

(22)申请日 2018.09.04

(71)申请人 中国平安人寿保险股份有限公司
地址 518048 广东省深圳市福田区益田路
5033号平安金融中心14、15、16、41、
44、45、46层

(72)发明人 陈石

(74)专利代理机构 广州华进联合专利商标代理
有限公司 44224

代理人 王宁

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

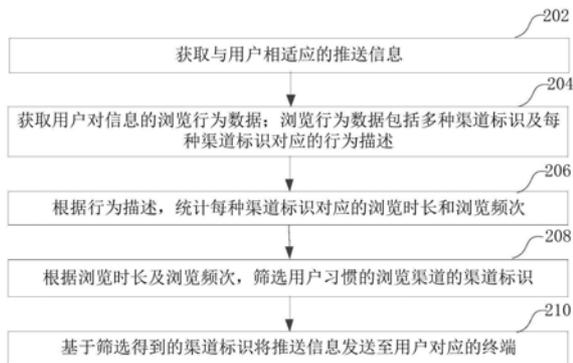
权利要求书2页 说明书11页 附图3页

(54)发明名称

信息推送方法、装置、计算机设备和存储介
质

(57)摘要

本申请涉及一种基于大数据分析的信息推
送方法、装置、计算机设备和存储介质。所述方法
包括：获取与用户相适应的待推送信息；获取所
述用户对信息的浏览行为数据；所述浏览行为数
据包括多种渠道标识及每种所述渠道标识对应
的行为描述；根据所述行为描述，统计每种所述
渠道标识对应的浏览时长和浏览频次；根据所述
浏览时长及浏览频次，筛选所述用户习惯的浏览
渠道的渠道标识；基于筛选得到的渠道标识将所
述待推送信息发送至所述用户对应的终端。采用
本方法能够减少服务器资源占用。



1. 一种信息推送方法,所述方法包括:
 - 获取与用户相适应的待推送信息;
 - 获取所述用户对信息的浏览行为数据;所述浏览行为数据包括多种渠道标识及每种所述渠道标识对应的行为描述;
 - 根据所述行为描述,统计每种所述渠道标识对应的浏览时长和浏览频次;
 - 根据所述浏览时长及浏览频次,筛选所述用户习惯的浏览渠道的渠道标识;
 - 基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。
2. 根据权利要求1所述的方法,其特征在于,所述获取与用户相适应的待推送信息,包括:获取所述用户的身份属性数据;
 - 根据所述身份属性数据,确定所述用户对应的用户类别;
 - 获取所述用户类别对应的待推荐产品列表,所述待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值;
 - 根据所述推荐热度值,将待推荐产品列表中一个或多个待推荐产品信息标记为与所述用户相适应的待推送信息。
3. 根据权利要求1所述的方法,其特征在于,所述获取用户对信息的浏览行为数据,包括:获取多个子模型,确定多个所述子模型分别对应的权值因子;
 - 根据多个所述子模型及分别对应的权值因子,生成第一分析模型;
 - 获取用户样本数据及对应的分类标签,将所述用户样本数据输入所述第一分析模型,得到中间分类结果;
 - 计算所述中间分类结果与所述分类标签的差异,根据所述差异对所述第一分析模型进行调整,得到第二分析模型;
 - 基于所述第二分析模型计算多个用户的潜在价值;
 - 获取所述潜在价值超过第一阈值的用户的浏览行为数据。
4. 根据权利要求1所述的方法,其特征在于,所述根据所述浏览时长及浏览频次,筛选所述用户习惯的浏览渠道的渠道标识,包括:
 - 获取多个用户对应的局域网信息;
 - 根据所述局域网信息,对多个所述用户进行群体划分,得到多个用户群组;
 - 确定每个用户群组对应的目标渠道标识,将所述目标渠道标识标记为用户群组中每个用户习惯的浏览渠道的渠道标识。
5. 根据权利要求1所述的方法,其特征在于,所述基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端,包括:
 - 获取所述用户的用户标签及对应的初始权重;
 - 计算所述用户标签对所述待推送信息的影响权重;
 - 根据所述初始权重和所述影响权重,计算用户对所述待推送信息的转化率;
 - 当所述转化率达到第二阈值时,基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。
6. 根据权利要求5所述的方法,其特征在于,所述当转化率达到第二阈值时,基于筛选得到的渠道标识将所述待推送信息发送至用户对应的终端,包括:
 - 当所述转化率达到第二阈值时,获取所述用户对应的社交网络图;所述社交网络图包

括目标用户节点和多个关联用户节点；

计算所述目标用户节点以及每个所述关联用户节点的潜在价值；

计算每个关联用户节点与所述用户节点的亲密度；

根据所述亲密度及所述潜在价值，计算所述用户的传递价值；

根据所述转化率及所述传递价值，计算用户对所述待推送信息的转化影响率；

当所述转化影响率达到第三阈值时，基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

7. 一种信息推送装置，其特征在于，所述装置包括：

内容筛选模块，用于获取与用户相适应的待推送信息；

渠道筛选模块，用于获取所述用户对信息的浏览行为数据；所述浏览行为数据包括多种渠道标识及每种所述渠道标识对应的行为描述；根据所述行为描述，统计每种所述渠道标识对应的浏览时长和浏览频次；根据所述浏览时长及浏览频次，筛选所述用户习惯的浏览渠道的渠道标识；

信息推送模块，用于基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

8. 根据权利要求7所述的装置，其特征在于，所述内容筛选模块还用于根据所述身份属性数据，确定所述用户对应的用户类别；获取所述用户类别对应的待推荐产品列表，所述待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值；根据所述推荐热度值，将待推荐产品列表中一个或多个待推荐产品信息标记为与所述用户相适应的待推送信息。

9. 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至6中任一项所述方法的步骤。

10. 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至6中任一项所述的方法的步骤。

信息推送方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及计算机技术领域,特别是涉及一种信息推送方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着信息时代的发展,企业常常通过电话、短信、邮件等渠道向用户推送产品信息或活动信息。然而,传统方式中大多企业对于信息推送渠道的选择主要是根据能够获取到的用户的联系方式。换言之,当企业获取到用户的多种联系方式时,可能会通过不同渠道多次向用户推送相同的信息,不仅造成信息冗余,也造成服务器资源的浪费。

发明内容

[0003] 基于此,有必要针对上述技术问题,提供一种能够减少服务器资源占用的信息推送方法、装置、计算机设备和存储介质。

[0004] 一种信息推送方法,所述方法包括:获取与用户相适应的待推送信息;获取所述用户对信息的浏览行为数据;所述浏览行为数据包括多种渠道标识及每种所述渠道标识对应的行为描述;根据所述行为描述,统计每种所述渠道标识对应的浏览时长和浏览频次;根据所述浏览时长及浏览频次,筛选所述用户习惯的浏览渠道的渠道标识;基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

[0005] 在其中一个实施例中,所述获取与用户相适应的待推送信息,包括:获取所述用户的身份属性数据;根据所述身份属性数据,确定所述用户对应的用户类别;获取所述用户类别对应的待推荐产品列表,所述待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值;根据所述推荐热度值,将待推荐产品列表中一个或多个待推荐产品信息标记为与所述用户相适应的待推送信息。

[0006] 在其中一个实施例中,所述获取用户对信息的浏览行为数据,包括:获取多个子模型,确定多个所述子模型分别对应的权值因子;根据多个所述子模型及分别对应的权值因子,生成第一分析模型;获取用户样本数据及对应的分类标签,将所述用户样本数据输入所述第一分析模型,得到中间分类结果;计算所述中间分类结果与所述分类标签的差异,根据所述差异对所述第一分析模型进行调整,得到第二分析模型;基于所述第二分析模型计算多个用户的潜在价值;获取所述潜在价值超过第一阈值的用户的浏览行为数据。

[0007] 在其中一个实施例中,所述根据所述浏览时长及浏览频次,筛选所述用户习惯的浏览渠道的渠道标识,包括:获取多个用户对应的局域网信息;根据所述局域网信息,对多个所述用户进行群体划分,得到多个用户群组;确定每个用户群组对应的目标渠道标识,将所述目标渠道标识标记为用户群组中每个用户习惯的浏览渠道的渠道标识。

[0008] 在其中一个实施例中,所述基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端,包括:获取所述用户的用户标签及对应的初始权重;计算所述用户标签对所述待推送信息的影响权重;根据所述初始权重和所述影响权重,计算用户对所述待推

送信息的转化率；当所述转化率达到第二阈值时，基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

[0009] 在其中一个实施例中，所述当转化率达到第二阈值时，基于筛选得到的渠道标识将所述待推送信息发送至用户对应的终端，包括：当所述转化率达到第二阈值时，获取所述用户对应的社交网络图；所述社交网络图包括目标用户节点和多个关联用户节点；计算所述目标用户节点以及每个所述关联用户节点的潜在价值；计算每个关联用户节点与所述用户节点的亲密度；根据所述亲密度及所述潜在价值，计算所述用户的传递价值；根据所述转化率及所述传递价值，计算用户对所述待推送信息的转化影响率；当所述转化影响率达到第三阈值时，基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

[0010] 一种信息推送装置，所述装置包括：内容筛选模块，用于获取与用户相适应的待推送信息；渠道筛选模块，用于获取所述用户对信息的浏览行为数据；所述浏览行为数据包括多种渠道标识及每种所述渠道标识对应的行为描述；根据所述行为描述，统计每种所述渠道标识对应的浏览时长和浏览频次；根据所述浏览时长及浏览频次，筛选所述用户习惯的浏览渠道的渠道标识；信息推送模块，用于基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。

[0011] 在其中一个实施例中，所述内容筛选模块还用于根据所述身份属性数据，确定所述用户对应的用户类别；获取所述用户类别对应的待推荐产品列表，所述待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值；根据所述推荐热度值，将待推荐产品列表中一个或多个待推荐产品信息标记为与所述用户相适应的待推送信息。

[0012] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现本申请任意一个实施例中提供的信息推送方法的步骤。

[0013] 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本申请任意一个实施例中提供的信息推送方法的步骤。

[0014] 上述信息推送方法、装置、计算机设备和存储介质，根据用户对信息的浏览行为数据，可以统计得到用户基于不同渠道进行信息浏览的浏览时长和浏览频次；根据所述浏览时长及浏览频次，可以筛选得到所述用户习惯的浏览渠道的渠道标识；当需要向用户推送某种信息时，获取与用户相适应的待推送信息，可以基于筛选得到的渠道标识将所述待推送信息发送至所述用户对应的终端。由于监测每个用户的习惯阅读方式，从多个信息浏览渠道中仅筛选一种渠道进行信息推送，避免了信息冗余，同时减少服务器资源占用；筛选得到的渠道为用户习惯的信息浏览渠道，且获取与用户相适应的待推送信息，从推送内容和推送渠道上更加贴合用户习惯，从而提高信息推送准确率，使得信息推送价值提高，进而间接提高服务器资源利用率。

附图说明

[0015] 图1为一个实施例中信息推送方法的应用场景图；

[0016] 图2为一个实施例中信息推送方法的流程示意图；

[0017] 图3为一个实施例中待推送信息确定步骤的流程示意图；

[0018] 图4为一个实施例中社交网络图的示意图；

[0019] 图5为一个实施例中信息推送装置的结构框图；

[0020] 图6为一个实施例中计算机设备的内部结构图。

具体实施方式

[0021] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0022] 本申请提供的信息推送方法,可以应用于如图1所示的应用环境中。其中,终端102与服务器104通过网络进行通信。其中,终端102可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。用户可以在终端102基于不同信息浏览渠道(以下简称“渠道”)进行信息浏览。渠道可以是邮件、电话、短信或应用程序等。服务器104收集用户基于不同渠道对信息的浏览行为数据。浏览行为数据包括多种渠道标识及每种渠道标识对应的行为描述。例如,用户基于电话渠道进行信息浏览时是否挂断、接通时长、业务员反馈的通话内容等;基于邮件渠道进行信息浏览时邮件读取时间点,读取时长等。服务器对海量用户的浏览行为数据进行统计分析,得到用户基于不同渠道对信息的浏览时长和浏览频次。服务器预设了多个指标区间及分别对应的分值,根据统计得到的不同渠道标识对应的浏览时长和浏览频次,计算不同渠道标识对应的分值,将分值最高的渠道标识标记为用户习惯的浏览渠道的渠道标识。当需要对该用户进行信息推送时,获取与该用户相适应的待推送信息,基于该用户习惯的浏览渠道将待推送信息推送至用户对应的终端。上述信息推送过程,监测每个用户的习惯阅读方式,从多个信息浏览渠道中筛选一个用户习惯的渠道进行信息推送,避免了信息冗余,同时减少服务器资源占用。

[0023] 在一个实施例中,如图2所示,提供了一种信息推送方法,以该方法应用于图1中的服务器为例进行说明,包括以下步骤:

[0024] 步骤202,获取与用户相适应的待推送信息。

[0025] 推荐信息可以是产品信息、也可以是新闻信息,还可以是其他信息。当推荐产品信息时,服务器对每个用户可能感兴趣的产品进行预测,进而根据预测结果确定最终的推送内容。

[0026] 步骤204,获取用户对信息的浏览行为数据;浏览行为数据包括多种渠道标识及每种渠道标识对应的行为描述。

[0027] 用户具有对应的标识字段。服务器从用户在企业留存的身份信息中提取基础标识字段。标识字段可以是用户的身份识别字段,如姓名、身份证号、手机号、邮箱账号、应用程序账号、常用设备信息等。常用设备信息可以是IMEI(International Mobile Equipment Identity,国际移动设备识别码)、IP地址、设备指纹、操作系统版本号、序列号等。

[0028] 不同互联网服务器上运行了不同网络平台。用户在使用各类设备基于不同渠道浏览待推送信息时,将在相应互联网服务器中留下浏览记录。浏览记录可以以日志或文件等的形式存储。服务器根据用户的基础标识字段,生成数据提取请求,将数据提取请求发送至互联网服务器。互联网服务器查找包含标识字段的浏览记录,将查找到的浏览记录返回至服务器。浏览记录是指用户基于移动终端、汽车、智能机器人、智能穿戴设备等发生的信息浏览行为(如浏览行为、评论行为、转发行为、收藏行为等)的行为数据。浏览行为数据包括

多种渠道标识及每种渠道标识对应的行为描述。行为描述包括对用户的信息浏览过程中浏览时间、是否评论、是否转发、是否收藏等细节信息。

[0029] 步骤206,根据行为描述,统计每种渠道标识对应的浏览时长和浏览频次。

[0030] 服务器对海量用户的浏览行为数据进行统计分析,得到每个用户基于不同渠道对信息的浏览时长和浏览频次。为了减少对服务器资源的占用,可以在数据库空闲时间对海量浏览行为数据进行统计分析。

[0031] 步骤208,根据浏览时长及浏览频次,筛选用户习惯的浏览渠道的渠道标识。

[0032] 服务器预设了多个指标区间及分别对应的分值。浏览时长和浏览频次分别作为一种指标具有对应的指标区间。服务器根据统计得到的不同渠道标识对应的浏览时长和浏览频次,计算不同渠道标识对应的分值,将分值最高的渠道标识标记为用户习惯的浏览渠道的渠道标识。

[0033] 在另一个实施例中,服务器还监测用户的习惯阅读时段(或信息禁发时段),在用户习惯阅读时段通过筛选得到的渠道进行信息推送。具体的,浏览行为数据还包括用户基于不同渠道浏览信息时的浏览时间。服务器统计用户分别在多个预设监控时段发生信息浏览行为的频次和时长,筛选发生信息浏览行为的频次最多和时长最长的监控时段,将筛选得到的监控时段作为用户的习惯阅读时段。在又一个实施例中,习惯阅读时段也可以由用户自由设置。

[0034] 步骤210,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0035] 当需要对该用户进行信息推送时,服务器获取与该用户相适应的待推送信息,基于该用户习惯的浏览渠道将待推送信息推送至用户对应的终端。

[0036] 本实施例中,根据用户对信息的浏览行为数据,可以统计得到用户基于不同渠道进行信息浏览的浏览时长和浏览频次;根据浏览时长及浏览频次,可以筛选得到用户习惯的浏览渠道的渠道标识;当需要向用户推送某种信息时,获取与用户相适应的待推送信息,可以基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。由于监测每个用户的习惯阅读方式,从多个信息浏览渠道中仅筛选一种渠道进行信息推送,避免了信息冗余,同时减少服务器资源占用;筛选得到的渠道为用户习惯的信息浏览渠道,且获取与用户相适应的待推送信息,从推送内容和推送渠道上更加贴合用户习惯,从而提高信息推送准确率,使得信息推送价值提高,进而间接提高服务器资源利用率。

[0037] 在一个实施例中,如图3所示,获取与用户相适应的待推送信息,即待推送信息确定的步骤,包括:

[0038] 步骤302,获取用户的身份属性数据。

[0039] 服务器获取当前的用户(以下简称“当前用户”)的身份属性数据。身份属性数据包括年龄、性别、婚姻状况、文化程度、职业、财产状况、历史购买记录、健康状况等。

[0040] 步骤304,根据身份属性数据,确定用户对应的用户类别。

[0041] 用户类别是指预先根据历史用户的身份属性数据划分的类别。不同用户类别所对应的感兴趣的产品信息不同。服务器根据当前用户的身份属性数据与历史用户的身份属性数据之间的相似度确定当前用户的用户类别。相似度的计算可以采用距离度量的方式,比如,可以采用欧式距离,也可以采用曼哈顿距离。具体的,服务器获取与当前用户身份属性数据对应的当前用户特征向量。当前用户特征向量是指将当前用户身份属性数据中各个指

标转换为可以衡量的数字表示的形式。比如,年龄可以直接用实际的年龄数字表示,性别分别采用0表示女,1表示男,婚姻状况可以分为未婚、已婚、离异三种状态,分别采用1、2和3表示。健康状况可以简单分为健康、不健康以及亚健康三种状态,当然也可以细分为0-100的健康数值。职业按照行业类别分为多个类,并设置每一类对应的数字。服务器根据当前用户身份属性数据对应的当前用户特征向量与历史用户身份信息属性数据对应的历史用户特征向量之间的距离计算相似度。

[0042] 在另一个实施例中,相似度也可以通过获取用户身份属性数据中每个元素与历史用户身份属性数据中相应元素之间的相似度,然后进行加权求和得到的。比如,年龄相似度可以是年龄的比值,且年龄大的作为分母,年龄小的作为分子,如20岁和18岁的年龄相似度为 $18/20=0.9$ 。

[0043] 步骤306,获取用户类别对应的待推荐产品列表,待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值。

[0044] 待推荐产品列表是与用户类别对应的,不同的用户类别对应不同的待推荐产品类别,待推荐产品列表中记录了每个待推荐产品信息对应的推荐热度值。推荐热度值反映了用户对待推荐产品信息的感兴趣程度,推荐热度值与用户的感兴趣程度成正相关,即推荐热度值越大,表明待推荐产品信息所对应的感兴趣程度越大。在一个实施例中,推荐热度值是动态变化的。

[0045] 步骤308,根据推荐热度值,将待推荐产品列表中一个或多个待推荐产品信息标记为与用户相适应的待推送信息。

[0046] 服务器将待推荐产品列表中推荐热度值最大的待推荐产品信息作为与当前用户相适应的待推送信息。在另一个实施例中,服务器将待推荐产品列表中的待推荐产品信息按照推荐热度值从大到小进行排序,获取前预设个数的待推荐产品信息作为待推送信息。

[0047] 本实施例中,通过确定当前用户的用户类别,然后根据目标用户类别对应的待推荐险种列表中待推荐险种信息的推荐热度值进行推荐,有针对性地根据用户类别进行推荐,提高了推荐的准确度,减少了推荐资源的浪费。

[0048] 在一个实施例中,获取用户对信息的浏览行为数据,包括:获取多个子模型,确定多个子模型分别对应的权值因子;根据多个子模型及分别对应的权值因子,生成第一分析模型;获取用户样本数据及对应的分类标签,将用户样本数据输入第一分析模型,得到中间分类结果;计算中间分类结果与分类标签的差异,根据差异对第一分析模型进行调整,得到第二分析模型;基于第二分析模型计算多个用户的潜在价值;获取潜在价值超过第一阈值的用户的浏览行为数据。

[0049] 服务器每隔一个时段构建相应时段周期(以下简称“周期”)的潜在价值分析模型。周期的时间长度可以根据需求自由设定,如1年。当前周期对应的潜在价值分析模型可以是利用多个历史周期的潜在价值分析模型构建而成。为了描述方便,将历史周期的潜在价值分析模型称作“子模型”。最初的子模型可以是服务器利用大量用户样本数据对初始模型训练得到。

[0050] 服务器获取多个历史周期的用户样本数据,针对每个用户的用户样本数据添加对应的品质标签。为了减少人工标注的繁琐,服务器根据用户样本数据建立用户画像,基于用户画像自动生成相应用户的品质标签。具体的,服务器对用户样本数据进行清洗等处理,得

到用户对应的多个属性标签,如用户的年龄、性别、职业、婚姻状况、文化程度、职业、财产保障、健康状况等。服务器将获取的多个属性标签组成为一个文本向量,将组成的文本向量作为该用户的用户画像。用户画像作为实际用户的虚拟代表,其往往是根据产品和市场来构建出来的,反应了真实用户的特征和需求。服务器预存储了多种属性标签组合以及每种组合分别对应的品质标签。服务器基于预存储的属性标签组合与品质标签的对应关系,对用户画像进行转化,得到相应用户标识对应的品质标签。品质标签可以是分值等定量指标,也可以是优、良、差等定性指标。

[0051] 服务器基于大量用户样本数据及对应的品质标签对初始模型训练,得到相应子模型。初始模型可以是特征分类模型和特征融合模型拟合得到。初始模型包括多个用户指标,每种用户指标具有对应的多种用户属性,如用户指标“性别”对应的用户属性可以是“男”或“女”。服务器计算各个用户指标对应的熵值增益。计算熵值增益的公式可以是:

[0052]

$$GA = -(M \times \log_2 M) + (1 - M) \log_2^{(1-M)} - \sum_{i=1}^{i=n} A_i \times (-a_i \times \log_2^{a_i} - (1 - a_i) \times \log_2^{1-a_i})。$$

[0053] 其中,GA表示计算的用户指标A的熵值增益;M表示用户潜在价值达到阈值的概率;A_i表示对应用户指标A的用户属性i的数量占用户样本数据中用户属性的总数量的比例,a_i表示用户属性i以用户指标A的数量为基数的用户品质潜在价值达到阈值的概率,n表示对应用户指标A的用户属性的个数。服务器将多个用户指标的熵值增益加权求和得到相应用户指标组合对应的熵值增益。服务器根据各个用户指标组合分别对应的熵值增益和品质标签,通过特征分类算法对第一预设模型进行训练,得到特征分类模型。特征分类算法可以是GBDT(Gradient Boost Decision Tree,梯度提升树算法)与(Logistic Regression,逻辑回归算法)的结合。

[0054] 服务器基于用户样本数据训练得到特征融合模型。具体的,若用户数据是通网络平台爬取得到的,不同网络平台对同一用户指标的命名方式可能不同,为了减少命名差异对模型训练的影响,服务器对各个用户指标进行同义扩展处理,得到各个用户指标组合分别对应的扩展指标组合。服务器分别获取用户指标中各个分词对应的同义词,将分词与对应的同义词形成扩展词语集合。每个分词都存在对应的扩展词语集合,如用户指标组合A为{a,b,c},则用户指标组合中的每个用户指标都存在对应的扩展词语集合,如用户指标a对应的扩展词语集合为{a,a1,a2}。服务器按照与用户指标组合中各个用户指标出现的顺序,从各个用户指标对应的扩展词语集合中任意选择一个词语,按顺序形成一个扩展指标集合。当从扩展词语集合中选择不同的词语时,则形成不同的扩展指标集合,不同的扩展指标集合组成扩展指标组合。服务器根据各个扩展指标组合及分别对应的价值分析结果,通过特征融合算法对第二预设模型进行训练,得到特征融合模型。特征融合算法可以是随机森林算法等。

[0055] 先形成每个用户指标对应的扩展词语集合,再通过扩展词语集合形成各个用户指标组合对应的扩展指标组合,大大提高了用户指标的扩展度,扩展后的各个用户指标表达了与原始的用户指标相同或相近的含义,提高了用户指标的有效覆盖范围,从而在后续输入已训练的特征融合模型后,可提高潜在价值分析精准性。

[0056] 服务器将特征分类模型与特征融合模型拟合,得到相应子模型。在一个具体的实

施例中,服务器对逻辑回归模型、GBDT (Gradient Boost Decision Tree,非线性模型)、(Logistic Regression,逻辑回归模型)、随机森林模型进行线性拟合,得到子模型。例如,子模型=逻辑回归模型*W1+GBDT*W2+LR*W3+随机森林模型*W4。其中,Wi为权重因子。不同类型模型存在ROC (receiver operating characteristic curve,受试者工作特征) 差异性,这里把不同类型模型拟合,能够提高用户潜在价值分析精准性。

[0057] 每个子模型具有对应时间标签。时间标签可以是根据子模型的构建时段生成,如2017,20170317等。服务器根据时间衰减函数 $e^{-\frac{\Delta t}{T}}$ 确定各个子模型的贡献率,即确定多个子模型分别对应的权值因子。其中, Δt 为时间标签与当前时间的的时间差; T 为最佳时间长度。例如,时间标签“2017”年与当前时间“2018”的时间差 $\Delta t=1$; T 可以是滑窗函数对应的时间跨度,即筛选得到的子模型的数量。容易理解,离当前时段越远的子模型是采用较早历史用户数据训练,用于分析当前时段用户潜在价值的参考意义(即贡献率)越小。换言之,时间差越大,对应子模型对品质分析的贡献率越小,从而可以基于时间衰减函数确定多个子模型的权值因子。

[0058] 服务器基于多个子模型以及分别对应的权值因子进行线性回归运算,得到第一分析模型。为了提高第一分析模型的准确性,服务器对第一分析模型进行训练强化。具体的,服务器获取当前周期内多个用户的用户样本数据。用户样本数据具有对应的分类标签。用户样本数据包括用户多个维度的信息,如年龄、职业、家庭成员等。服务器将当前周期的用户样本数据输入第一分析模型,得到中间分类结果。服务器计算中间分类结果与分类标签的差异,根据差异对第一分析模型进行调整,得到第二分析模型。

[0059] 服务器将当前用户的身份属性数据输入第二分析模型,得到当前用户的潜在价值。服务器对潜在价值达到第一阈值的用户进行信息推送。

[0060] 本实施例中,根据用户的潜在价值判断是否有必要向该用户进行信息推送,可以提高了推荐的准确度,减少了推荐资源的浪费。由于利用用户样本数据预先构建潜在价值分析模型,基于分析模型只需将用户数据作为入参即可快速得到相应用户的潜在价值,也可以综合考虑用户多个维度的信息,从而不仅提高用户潜在价值分析效率,也提高了用户潜在价值分析准确率。

[0061] 在一个实施例中,根据浏览时长及浏览频次,筛选用户习惯的浏览渠道的渠道标识,包括:获取多个用户对应的局域网信息;根据局域网信息,对多个用户进行群体划分,得到多个用户群组;确定每个用户群组对应的目标渠道标识,将目标渠道标识标记为用户群组中每个用户习惯的浏览渠道的渠道标识。

[0062] 同一企业的用户常常习惯采用相同或相似的信息浏览渠道,如基于工作邮箱浏览待推送信息。服务器根据用户所在局域网对用户进行群体划分,以用户群组为单位进行浏览渠道偏好的统计。具体的,浏览行为数据还包括局域网标识。服务器将包括同一个局域网标识的多个用户标识统计为一个用户群组。服务器筛选每个用户群组中不同用户分别对应的习惯渠道标识。习惯渠道标识对应的浏览时长及浏览频次分别达到相应属性值。根据习惯渠道标识的数量,确定每个用户群组对应的目标渠道标识,将目标渠道标识标记为用户群组中每个用户习惯的浏览渠道的渠道标识。

[0063] 本实施例中,对大量用户进行群体划分,以用户群组为单位进行用户习惯渠道分

析,可以避免服务器对大量用户逐一分析的繁琐,简化信息推送渠道选择步骤。

[0064] 在一个实施例中,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端,包括:获取用户的用户标签及对应的初始权重;计算用户标签对待推送信息的影响权重;根据初始权重和影响权重,计算用户对待推送信息的转化率;当转化率达到第二阈值时,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0065] 服务器根据用户的身份属性数据确定用户对应的用户标签,如性别、年龄等。一个用户往往对应多个用户标签,而用户对不同事物的感兴趣度不同,所以每个用户标签所占的权重也是不同的。服务器获取用户标签的初始权重,也就是获取用户的具体特征。

[0066] 服务器计算用户标签对待推送信息的影响权重。具体的,对于不同的待推送信息,每个用户标签的影响权重不同。比如,如果推荐信息为高跟鞋信息,那么用户标签中的女性、购物标签所占的权重大一些。计算用户标签的影响权重可以采用TF-idf (term frequency-inverse document frequency,词频-逆向文件频率)算法。词频指的是某一个给定的词语在该文件中出现的频率,这个数字是对词数的归一化,以防止它偏向长的文件。逆向文件频率是一个词语普遍重要性的度量。某一特定文件内的高词语频率,以及该词语在整个文件集合中的低文件频率可以产生高权重的TF-idf值。

[0067] 服务器根据初始权重和影响权重计算用户对待推送信息的转化率。用户对待推送信息的转化率是指用户对推送内容关注或者购买的可能概率。具体的,用户标签的初始权重为 $Weight(U, T_i)$,其中 U 代表用户, T_i 代表标签, i 表示用户的第 i 种标签。计算得到的用户标签对待推送信息的影响权重为 $W(T_i, C)$,其中, C 代表待推送信息。用户标签的初始权重 $Weight(U, T_i)$ 和用户标签对待推送信息的影响权重 $W(T_i, C)$ 构成了用户 U 的特征向量,具体表示为:用户 U 的特征向量= $[Weight(U, T_0) * W(T_0, C), Weight(U, T_1) * W(T_1, C), Weight(U, T_2) * W(T_2, C), \dots]$,根据已有的样本数据即之前大量具有相同或相似的用户标签的用户是否关注或购买待推送信息作为模板进行机器训练,用训练出的模型,根据用户的特征向量进行回归学习得到用户对待推送信息的转化率。服务器对转化率达到第二阈值的用户进行信息推送。

[0068] 本实施例中,根据用户的转化率判断是否有必要向该用户进行信息推送,可以提高推荐的准确度,减少了推荐资源的浪费。

[0069] 在一个实施例中,当转化率达到第二阈值时,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端,包括:当转化率达到第二阈值时,获取用户对应的社交网络图;社交网络图包括目标用户节点和多个关联用户节点;计算目标用户节点以及每个关联用户节点的潜在价值;计算每个关联用户节点与用户节点的亲密度;根据亲密度及潜在价值,计算用户的传递价值;根据转化率及传递价值,计算用户对待推送信息的转化影响率;当转化影响率达到第三阈值时,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0070] 为了进一步提高信息推送准确率,当用户对待推送信息的转化率达到第二阈值时,服务器进一步计算用户对待推送信息的转化影响率。转化影响率是指用户对社交关系中的其他用户在待推送信息上的影响度,即用户对待推送信息关注或者购买后,其社交关系中的其他用户看到该用户关注或者购买后,也进行关注或者购买的概率。具体的,服务器获取用户对应的社交网络图。社交网络图是根据用户的社交关系数据生成的。社交关系数

据可以从预先指定的社交网站中爬取得到的。社交网站中的社交关系可以是好友关系、相互关注等彼此之间互相关联的关系。社交关系还包括用户动作相关联的数据,比如,用户发布或者分享信息影响到好友进行评论、点赞、转发等。

[0071] 社交网络图包括多个用户节点以及用于连接用户节点的边线。具体的,服务器根据用户基础数据确定与用户标识对应的目标用户节点。用户基础数据包括用户自身的属性信息,如姓名、年龄、职业等。服务器根据用户关联数据确定每个关键用户对应的关联用户节点。用户关联数据包括与用户存在社交关系的其他用户的属性信息。如图4所示,服务器通过节点之间的边线来标识用户之间的关系,即根据用户之间的社交关系连接目标用户节点 v_1 与关联用户节点 v_i ($i \geq 2$),得到用户标识对应的社交网络图。

[0072] 根据社交网络图,服务器获取每个关联用户节点对应的身份属性数据,将身份属性数据输入预设的价值分析模型,得到每个关联用户节点对应的潜在价值。服务器利用预设计算公式计算每个关联用户节点与目标用户节点的亲密度。亲密度计算公式可以是:

$$[0073] \quad Q(v, w) = \frac{|N(v) \cap N(w)| + 1}{|N(v) \cup N(w)|}$$

[0074] 其中, $Q(v, w)$ 为关联用户节点 w 与目标用户节点 v 之间的亲密度; $N(v)$ 表示目标用户节点 v 的邻接节点集合;目标用户节点 v 和关联用户节点 w 的共同邻接节点数为 $|N(v) \cap N(w)|$;目标用户节点 v 和关联用户节点均不邻接的节点数为 $|N(v) \cup N(w)|$ 。

[0075] 服务器根据计算得到亲密度以及潜在价值,计算用户对应的传递价值。传递价值是指用户将待推送信息转告至其他用户,且其他用户对待推送信息感兴趣的概率。具体的,服务器筛选目标用户节点亲密度及潜在价值分别达到相应属性值的关联用户节点。服务器利用目标用户节点与筛选得到的关联用户节点构建目标群组,将目标群组作为用户的有效社交群组。服务器计算有效社交群组中多个群组成员的平均潜在价值,即计算目标用户节点对应潜在价值与多个关联用户节点对应潜在价值的平均值。服务器计算有效社交群组的群组规模及群组核度。群组规模是指群组成员的数量。群组核度是指群组成员中每个群组成员邻接组内其他群组成员的数量最小值。群组核度可以反映关联用户节点与关联用户节点之间的社交热度。服务器根据群组规模、群组核度及平均潜在价值确定用户标识对应的传递价值。

[0076] 服务器根据用户对待推送信息的转化率及传递价值,计算用户对待推送信息的转化影响率。服务器对转化影响率达到第三阈值的用户进行信息推送。

[0077] 本实施例中,不但考虑了推送的个体效应即用户对待推送信息的转化率,还考虑了推送的群体效应,即用户对待推送信息在社交关系中的影响度,也就是说,待推送信息的投放不仅是针对用户个人,更是针对用户的社交圈,扩大了推送的范围,降低了推送的成本。

[0078] 应该理解的是,虽然图2和图3的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图2和图3中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些子步骤或者阶段的执行顺序也不必然

是依次进行,而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0079] 在一个实施例中,如图5所示,提供了一种信息推送装置,包括:内容筛选模块502、渠道筛选模块504和信息推送模块506,其中:

[0080] 内容筛选模块502,用于获取与用户相适应的待推送信息;

[0081] 渠道筛选模块504,用于获取用户对信息的浏览行为数据;浏览行为数据包括多种渠道标识及每种渠道标识对应的行为描述;根据行为描述,统计每种渠道标识对应的浏览时长和浏览频次;根据浏览时长及浏览频次,筛选用户习惯的浏览渠道的渠道标识;

[0082] 信息推送模块506,用于基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0083] 在一个实施例中,内容筛选模块502还用于根据身份属性数据,确定用户对应的用户类别;获取用户类别对应的待推荐产品列表,待推荐产品列表中记录了多个待推荐产品信息对应的推荐热度值;根据推荐热度值,将待推荐产品列表中一个或多个待推荐产品信息标记为与用户相适应的待推送信息。

[0084] 在一个实施例中,渠道筛选模块504包括价值分析模块5042,用于获取多个子模型,确定多个子模型分别对应的权值因子;根据多个子模型及分别对应的权值因子,生成第一分析模型;获取用户样本数据及对应的分类标签,将用户样本数据输入第一分析模型,得到中间分类结果;计算中间分类结果与分类标签的差异,根据差异对第一分析模型进行调整,得到第二分析模型;基于第二分析模型计算多个用户的潜在价值;获取潜在价值超过第一阈值的用户的浏览行为数据。

[0085] 在一个实施例中,渠道筛选模块504还包括群组划分模块5044,用于获取多个用户对应的局域网信息;根据局域网信息,对多个用户进行群体划分,得到多个用户群组;确定每个用户群组对应的目标渠道标识,将目标渠道标识标记为用户群组中每个用户习惯的浏览渠道的渠道标识。

[0086] 在一个实施例中,信息推送模块506包括转化预测模块5062,用于获取用户的用户标签及对应的初始权重;计算用户标签对待推送信息的影响权重;根据初始权重和影响权重,计算用户对待推送信息的转化率;当转化率达到第二阈值时,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0087] 在一个实施例中,信息推送模块506还包括影响预测模块5064,用于当转化率达到第二阈值时,获取用户对应的社交网络图;社交网络图包括目标用户节点和多个关联用户节点;计算目标用户节点以及每个关联用户节点的潜在价值;计算每个关联用户节点与用户节点的亲密度;根据亲密度及潜在价值,计算用户的传递价值;根据转化率及传递价值,计算用户对待推送信息的转化影响率;当转化影响率达到第三阈值时,基于筛选得到的渠道标识将待推送信息发送至用户对应的终端。

[0088] 关于信息推送装置的具体限定可以参见上文中对于信息推送方法的限定,在此不再赘述。上述信息推送装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0089] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结

构图可以如图6所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储用户的身份属性数据和浏览行为数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种信息推送方法。

[0090] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0091] 一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现本申请任意一个实施例中提供的信息推送方法的步骤。

[0092] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0093] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0094] 以上实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

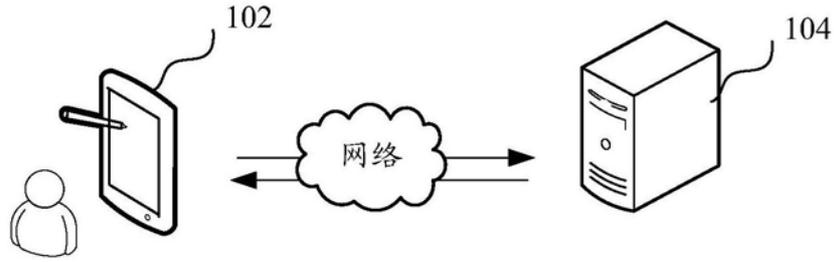


图1

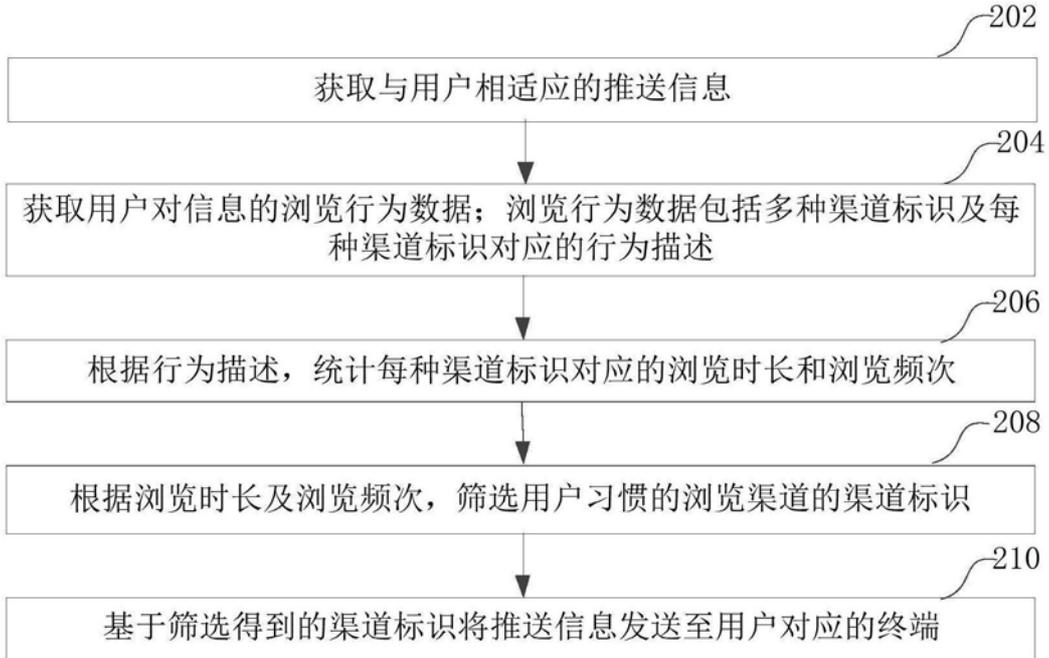


图2

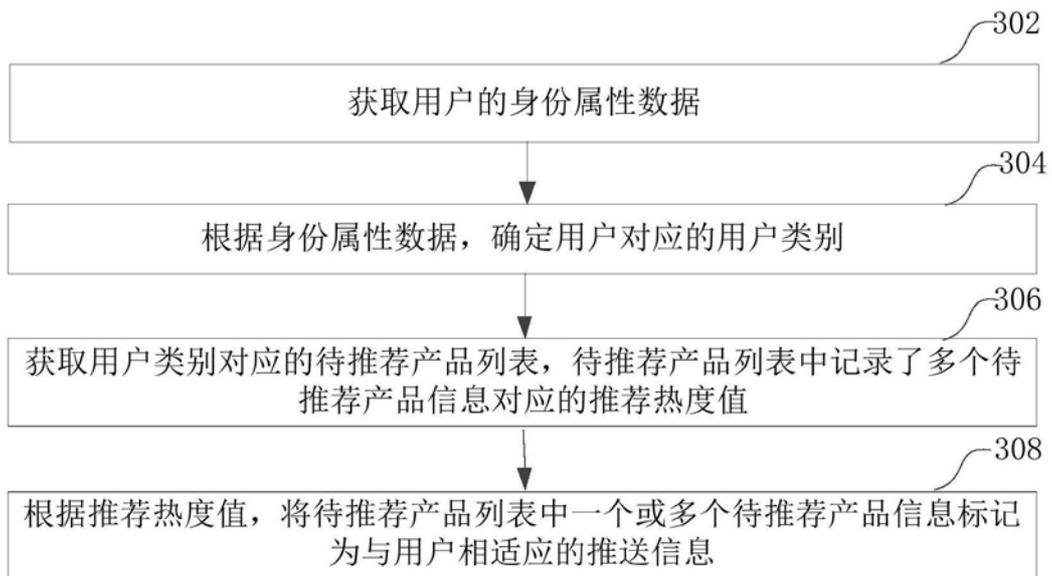


图3

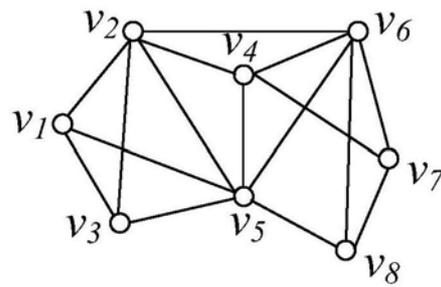


图4

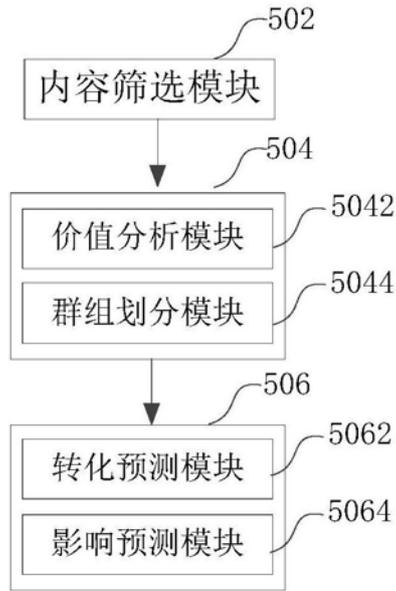


图5

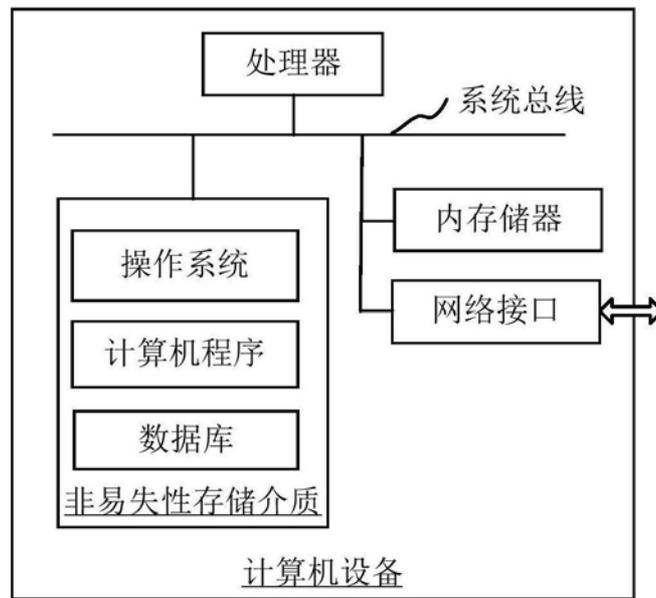


图6