



# (12)发明专利申请

(10)申请公布号 CN 109446431 A

(43)申请公布日 2019.03.08

(21)申请号 201811507489.6

(22)申请日 2018.12.10

(71)申请人 网易传媒科技(北京)有限公司  
地址 100084 北京市海淀区西北旺东路10  
号院中关村软件园西区7号楼A座5层

(72)发明人 高畅 刘洪彬 童志成 王彦争  
谢鹏 魏望

(74)专利代理机构 中科专利商标代理有限责任  
公司 11021  
代理人 吕雁葭

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

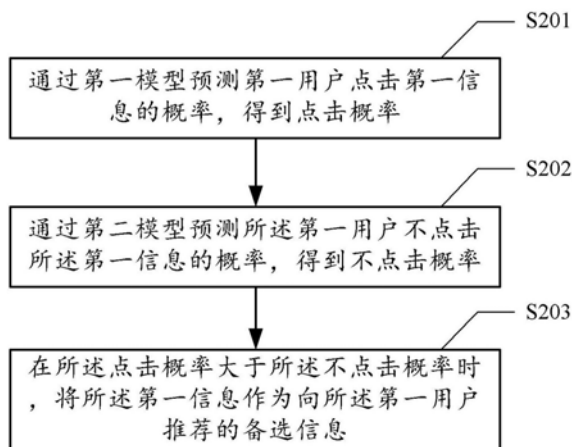
权利要求书2页 说明书11页 附图5页

## (54)发明名称

用于信息推荐的方法、装置、介质、和计算设备

## (57)摘要

本发明的实施方式提供了一种用于信息推荐的方法。所述方法包括通过第一模型预测第一用户点击第一信息的概率,得到点击概率,通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率,以及在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息。其中,所述第一模型与所述第二模型不同。本发明的方法使得筛选获得的用于信息推荐的备选信息更具有针对性,能够降低根据经验确定的推荐结果的不可控性,有效提高推荐结果的转化率。此外,本发明的实施方式提供了一种用于信息推荐的装置、介质和计算设备。



1. 一种用于信息推荐的方法,包括:  
通过第一模型预测第一用户点击第一信息的概率,得到点击概率;  
通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率;以及  
在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的  
备选信息;

其中:

所述第一模型与所述第二模型不同。

2. 根据权利要求1所述的方法,其中:

所述第一模型为用于预测用户点击信息的行为的模型;以及

所述第二模型为用于预测用户不点击信息的行为的模型。

3. 根据权利要求1所述的方法,还包括:

通过机器学习获得所述第一模型和所述第二模型,包括:

通过第一类特征数据训练所述第一模型;以及

通过第二类特征数据训练所述第二模型;

其中:

所述第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据;

所述第二类特征数据包括每个用户的不感兴趣标签、以及所述表征该每个用户的浏览行为的数据。

4. 根据权利要求3所述的方法,其中,所述浏览行为包括以下任意一种或多种行为:

点击信息的行为、评价信息的行为、浏览标题但未点击信息的行为、取消关注信息的行为、以及强制关闭信息的行为。

5. 根据权利要求4所述的方法,还包括:

获取原始数据,所述原始数据包括特定时间内预定网络平台中多个用户的与所述浏览行为相关的数据;以及

从所述原始数据中提取所述第一类特征数据和所述第二类特征数据。

6. 一种用于信息推荐的装置,包括:

点击概率预测模块,用于通过第一模型预测第一用户点击第一信息的概率,得到点击概率;

不点击概率预测模块,用于通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率;以及

决策模块,用于在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息;

其中:

所述第一模型与所述第二模型不同。

7. 根据权利要求6所述的装置,还包括:

模型获得模块,用于通过机器学习获得所述第一模型和所述第二模型,包括:

第一训练子模块,用于通过第一类特征数据训练所述第一模型;以及

第二训练子模块,用于通过第二类特征数据训练所述第二模型;

其中：

所述第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据；

所述第二类特征数据包括每个用户的不感兴趣标签、以及所述表征该每个用户的浏览行为的数据。

8. 根据权利要求7所述的装置,还包括：

原始数据获取模块,用于获取原始数据,所述原始数据包括特定时间内预定网络平台中多个用户的与所述浏览行为相关的数据;以及

特征数据提取模块,用于从所述原始数据中提取所述第一类特征数据和所述第二类特征数据。

9. 一种计算机可读存储介质,其上存储有可执行指令,所述指令在被处理器执行时使处理器执行根据权利要求1~5任意一项所述的方法。

10. 一种计算设备,包括：

一个或多个存储器,存储有可执行指令；

一个或多个处理器,执行所述可执行指令,用以实现根据权利要求1~5任意一项所述的方法。

## 用于信息推荐的方法、装置、介质、和计算设备

### 技术领域

[0001] 本发明的实施方式涉及互联网领域,更具体地,本发明的实施方式涉及一种用于信息推荐的方法、装置、介质和计算设备。

### 背景技术

[0002] 本部分旨在为权利要求书中陈述的本发明的实施方式提供背景或上下文。此处的描述不因为包括在本部分中就承认是现有技术。

[0003] 互联网的出现和普及给用户带来了大量的信息,满足了用户在信息时代对信息的需求,但随着网络的迅速发展而带来的网上信息量的大幅增长,使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息,对信息的使用效率反而降低了,这就是所谓的信息超载(information overload)问题。目前,针对信息过载问题的一种非常有潜力的办法是通过个性化推荐系统进行信息推荐。个性化推荐系统能够根据用户的信息需求、兴趣等,将用户感兴趣的信息、产品等个性化信息推荐给用户。一个好的推荐系统不仅能为用户提供个性化的服务,还能和用户之间建立密切关系,让用户对推荐产生依赖。

[0004] 推荐系统进行信息推荐的过程大致包括召回和排序两个阶段。召回主要完成目标集合的初步筛选,将可推荐的信息集合限定在一定范围内。排序主要完成对召回结果的精细化筛选,用于输出更为准确的推荐结果。

[0005] 现有技术中,无论是在召回阶段或者排序阶段(尤其是在召回阶段),进行信息筛选时的一种方案是预测出某一信息的点击概率,然后将该点击概率与固定阈值相比较,得到筛选结果。这种方案的缺点是这个固定阈值往往是通过人工先验确定的,缺乏理论根据,导致最终的计算精度非常容易受此影响。另一种方案是是对大量信息的预测点击概率进行排序,然后选取前N名信息。然而,这种方案中N值的确定依然是通过先验制定的,并没有从根本上解决由先验经验导致的推荐结果不可控的问题。

### 发明内容

[0006] 因此在现有技术中,在进行信息推荐时,如何更科学合理地筛选信息,是非常令人烦恼的过程。

[0007] 为此,非常需要一种更为科学合理的信息推荐机制,以使信息推荐的更具有针对性。

[0008] 在本上下文中,本发明的实施方式期望提供一种用于信息推荐的方法、装置、介质和计算设备。

[0009] 在本发明实施方式的第一方面中,提供了一种用于信息推荐的方法。所述方法包括:通过第一模型预测第一用户点击第一信息的概率,得到点击概率;通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率;以及在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息。其中,所述第一模型与所述第二模型不同。

[0010] 在本发明的一个实施例中,所述第一模型为用于预测用户点击信息的行为的模型,以及所述第二模型为用于预测用户不点击信息的行为的模型。

[0011] 在本发明的另一个实施例中,所述方法还包括通过机器学习获得所述第一模型和所述第二模型。具体包括通过第一类特征数据训练所述第一模型,以及通过第二类特征数据训练所述第二模型。其中,所述第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据,所述第二类特征数据包括每个用户的不感兴趣标签、以及所述表征该每个用户的浏览行为的数据。在本发明的一个实施例中,所述浏览行为包括以下任意一种或多种行为:点击信息的行为、评价信息的行为、浏览标题但未点击信息的行为、取消关注信息的行为、以及强制关闭信息的行为。

[0012] 在本发明的又一个实施例中,所述方法还包括:获取原始数据,所述原始数据包括特定时间内预定网络平台中多个用户的与所述浏览行为相关的数据;以及从所述原始数据中提取所述第一类特征数据和所述第二类特征数据。

[0013] 在本发明实施方式的第二方面中,提供了一种用于信息推荐的装置。所述装置包括点击概率预测模块、不点击概率预测模块以及决策模块。点击概率预测模块用于通过第一模型预测第一用户点击第一信息的概率,得到点击概率。不点击概率预测模块用于通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率。决策模块用于在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息。其中,所述第一模型与所述第二模型不同。

[0014] 在本发明的一个实施例中,所述第一模型为用于预测用户点击信息的行为的模型,以及所述第二模型为用于预测用户不点击信息的行为的模型。

[0015] 在本发明的另一个实施例中,所述装置还包括模型获得模块。模型获得模块用于通过机器学习获得所述第一模型和所述第二模型。该模型获取模块包括第一训练子模块以及第二训练子模块。第一训练子模块用于通过第一类特征数据训练所述第一模型。第二训练子模块用于通过第二类特征数据训练所述第二模型。其中,所述第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据,所述第二类特征数据包括每个用户的不感兴趣标签、以及所述表征该每个用户的浏览行为的数据。在本发明的一个实施例中,所述浏览行为包括以下任意一种或多种行为:点击信息的行为、评价信息的行为、浏览标题但未点击信息的行为、取消关注信息的行为、以及强制关闭信息的行为。

[0016] 在本发明的又一个实施例中,所述装置还包括原始数据获取模块和特征数据提取模块。原始数据获取模块用于获取原始数据,所述原始数据包括特定时间内预定网络平台中多个用户的与所述浏览行为相关的数据特征数据提取模块用于从所述原始数据中提取所述第一类特征数据和所述第二类特征数据。

[0017] 在本发明实施方式的第三方中,提供了一种计算机可读存储介质,其上存储有可执行指令,所述指令在被处理器执行时使处理器执行如上所述的用于信息推荐的方法。

[0018] 在本发明实施方式的第四方面中,提供了一种计算设备。所述计算设备包括存储有可执行指令的一个或多个存储器,以及一个或多个处理器。所述处理器执行所述可执行指令,用以实现如上所述的用于信息推荐的方法。

[0019] 根据本发明实施方式的方法、装置、介质和计算设备,在筛选用于信息推荐的备选信息时,更加科学,更具有针对性,能够降低现有技术中根据经验筛选的不可控性,能够有

效提高推荐信息的转化率。

### 附图说明

[0020] 通过参考附图阅读下文的详细描述,本发明示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本发明的若干实施方式,其中:

[0021] 图1示意性地示出了根据本发明实施方式的用于信息推荐的方法、装置、介质和计算设备的应用场景;

[0022] 图2示意性地示出了根据本发明一实施例的用于信息推荐的方法流程图;

[0023] 图3示意性地示出了根据本发明另一实施例的用于信息推荐的方法流程图;

[0024] 图4示意性地示出了根据本发明一实施例的通过机器学习获得第一模型和第二模型的流程图;

[0025] 图5示意性地示出了根据本发明实施例的用于信息推荐方法的构思示意图;

[0026] 图6示意性地示出了根据本发明实施例的用于信息推荐的装置的框图;

[0027] 图7示意性地示出了根据本发明实施例的用于信息推荐的程序产品的示意图;以及

[0028] 图8示意性地示出了根据本发明实施例的用于信息推荐的计算设备的示意图。

[0029] 在附图中,相同或对应的标号表示相同或对应的部分。

### 具体实施方式

[0030] 下面将参考若干示例性实施方式来描述本发明的原理和精神。应当理解,给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本发明,而并非以任何方式限制本发明的范围。相反,提供这些实施方式是为了使本公开更加透彻和完整,并且能够将本公开的范围完整地传达给本领域的技术人员。

[0031] 本领域技术人员知道,本发明的实施方式可以实现为一种系统、装置、设备、方法或计算机程序产品。因此,本公开可以具体实现为以下形式,即:完全的硬件、完全的软件(包括固件、驻留软件、微代码等),或者硬件和软件结合的形式。

[0032] 根据本发明的实施方式,提出了一种用于信息推荐的方法、介质、装置和计算设备。

[0033] 在本文中,附图中的任何元素数量均用于示例而非限制,以及任何命名都仅用于区分,而不具有任何限制含义。

[0034] 下面参考本发明的若干代表性实施方式,详细阐释本发明的原理和精神。

#### [0035] 发明概述

[0036] 本发明人发现,人们通常只有当做一件事情的动力大于阻力时,才会去做这件事情。同样的道理,对于同一信息可以认为只有在用户点击该信息的愿望大于不点击该信息的愿望时,用户才会真正点击这个信息。由此发明人联想到在信息推荐领域,如果用户点击一个信息的概率大于用户不点击该信息的概率时,就可以认为用户会点击该信息。按照这一逻辑,对于一条信息,如果能够预测得到用户点击该信息的点击概率和用户不点击该信息的不点击概率,然后根据该点击概率和不点击概率的比较来进行筛选,就能够使得筛选

出的用于推荐的信息的针对性更强、转化率更高。

[0037] 在介绍了本发明的基本原理之后,下面具体介绍本发明的各种非限制性实施方式。

[0038] 应用场景总览

[0039] 首先参考图1。

[0040] 图1示意性地示出了根据本发明实施方式的用于信息推荐的方法、装置、介质和计算设备的应用场景。

[0041] 如图1所示,该应用场景包括用户终端11和服务器12。用户终端11和服务器12通过网络连接。该网络可以为局域网、广域网、移动互联网等,可以包括各种连接类型,例如有线、无线通信链路等等。

[0042] 用户终端11可以是但不限于便携设备(例如:手机、平板、笔记本电脑等),也可以为个人电脑(PC,Personal Computer)。用户可以使用用户终端11通过网络与服务器12交互,以接收或发送消息等。用户终端11上可以安装有各种通讯客户端应用,例如购物类应用、网页浏览器应用、新闻类应用、即时通信工具、邮箱客户端、社交平台软件等(仅为示例)。

[0043] 服务器12可以是提供各种服务的服务器,例如对用户利用用户终端11所浏览的网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的用户请求等数据进行分析等处理,并将处理结果反馈给用户终端。例如,根据本发明的实施例,服务器12可以向用户终端11进行个性化信息推荐。

[0044] 需要说明的是,本发明实施例所提供的用于信息推荐的方法可以由服务器12执行。相应地,本发明实施例所提供的用于信息推荐的装置、介质或计算设备一般可以设置于服务器12中。或者,本发明实施例所提供的用于信息推荐的方法也可以由不同于服务器12且能够与用户终端11和/或服务器12通信的其他网站服务器或网站服务器集群执行。相应地,本发明实施例所提供的用于信息推荐的装置、介质或计算设备也可以设置于不同于服务器12且能够与用户终端11和/或服务器12通信的其他网站服务器或网站服务器集群中。

[0045] 应该理解,图1中的用户终端、网络和网站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用户终端、网络和服务器。

[0046] 示例性方法

[0047] 下面结合图1的应用场景,参考图2~图5来描述根据本发明示例性实施方式的用于信息推荐的方法。需要注意的是,上述应用场景仅是为了便于理解本发明的精神和原理而示出,本发明的实施方式在此方面不受任何限制。相反,本发明的实施方式可以应用于适用的任何场景。

[0048] 图2示意性地示出了根据本发明一实施例的用于信息推荐的方法流程图。

[0049] 如图2所示,该用于信息推荐的方法可以包括操作S201~操作S203。

[0050] 在操作S201,通过第一模型预测第一用户点击第一信息的概率,得到点击概率。

[0051] 在操作S202,通过第二模型预测该第一用户不点击该第一信息的概率,得到不点击概率,其中该第一模型与第二模型不同。

[0052] 然后,在操作S203,在该点击概率大于该不点击概率时,将该第一信息作为向该第一用户推荐的备选信息。

[0053] 根据本发明的实施例,该第一模型与第二模型不同,具体可以是第一模型与第二模型为两个不同的模型,其中,二者可以是通过不同的训练数据进行训练得到的。根据本发明的实施例,该第一模型为用于预测用户点击信息的行为的模型,以及该第二模型为用于预测用户不点击信息的行为的模型。

[0054] 根据本发明实施例的方法,与现有技术中根据先验经验(例如根据经验确定的固定阈值或者排序的前N个)来筛选推荐信息的方法相比,可以降低使用先验知识带来的结果的不可控性,提高信息推荐的针对性和有效性。

[0055] 图3示意性地示出了根据本发明另一实施例的用于信息推荐的方法流程图。

[0056] 如图3所示,根据本发明的另一实施例,该用于信息推荐的方法除了操作S201~操作S203以外,还可以包括操作S301~S303。

[0057] 首先,在操作S301,获取原始数据。该原始数据包括特定时间内预定网络平台中多个用户的与浏览行为相关的数据。

[0058] 根据本发明实施例,该浏览行为包括以下任意一种或多种行为:点击信息的行为、评价信息的行为、浏览标题但未点击信息的行为、取消关注信息的行为、以及强制关闭信息的行为等。

[0059] 表1示意性示出了根据本发明实施例的原始数据的一个实例。

[0060] 表1



[0061]

用户 ID	年龄	兴趣标签 (与用户喜欢的 东西相关)	不感兴趣标签 (与用户讨厌的 东西相关)	文章 ID	浏览行为	
					是否点击	其他数据
U1	12	体育, 科技	人文	A11	点击	好评
				A12	不点击	投诉
				...	...	...
U2	35	天文, 地理	历史, 财经	A21	点击	—
				A22	不点击	取消关注
				...	...	...
U3	36	软件	财经, 法律	A31	不点击	强制关闭
				A32	点击	收藏
				...	...	...
U4	27	专利	音乐, 美术	A41	不点击	浏览标题
				A42	点击	差评
				...	...	...
...	...	...	...	...	...	...

[0062] 如表1所示,该原始数据例如可以是在特定时间内从预定网络平台中搜集整理的大量用户(例如,U1、U2、U3、U4等用户)的浏览行为相关的数据。该原始数据可以包括用户的基本信息、以及用户的浏览行为数据。

[0063] 参考表1,用户的基本信息例如可以包括用户ID、年龄、用户的兴趣标签和不感兴趣标签。在一些实施例中,用户的基本信息例如可以包括用户的性别、登陆该预定网络平台时地理位置、时间信息、和/或使用的设备等信息。

[0064] 根据本发明的实施例,用户的兴趣标签,例如可以是用户(例如,用户U1)在该网络平台的个人信息中选择或填写的兴趣标签,也可以是根据用户U1的浏览行为统计获得的兴趣标签。具体而言,根据用户U1的浏览行为统计获得的兴趣标签,例如可以是对用户U1在该特定期内点击、或进行好评、或收藏的信息中所包含的标签的出现次数或频率进行统计,将出现次数或者频率最高的若干个标签作为用户U1的兴趣标签。

[0065] 根据本发明的实施例,用户的不感兴趣标签,例如可以是用户(例如,用户U1)在该网络平台的个人信息中选择或填写的不感兴趣标签,也可以是根据用户U1的浏览行为统计

获得的不感兴趣标签。具体而言,根据用户U1的浏览行为统计获得的不感兴趣标签,例如可以是对用户U1在该特定时期内的投诉的信息、或者给予差评的信息、或者强制关闭的信息中所包含的标签出现次数进行统计,并将其中出现次数或者频率最高的若干个标签若为用户U1的不感兴趣标签。

[0066] 继续参考表1,用户的浏览行为数据,可以包括例如用户对哪些文章进行过点击、以及对哪些曝光于用户的浏览界面的文章并未点击。在一些实施例中,如表1所示用户的浏览行为数据还可以包括其他数据,例如用户对哪些文章进行过评价以及评价内容、对哪些文章进行了收藏、对哪些文章进行取消关注、和/或对哪些文章进行强制关闭等。在表1中每个文章ID用于在该预定网络平台上唯一标识对应的文章内容,每个文章内容可以具有至少一个或多个反映该文章内容的标签。

[0067] 然后在操作S302,从该原始数据中提取该第一类特征数据和该第二类特征数据。例如具体可以是首先将原始数据规范化和结构化处理,然后从中提取该第一类特征数据和该第二类特征数据。

[0068] 根据本发明的实施例,该第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据。该第二类特征数据包括每个用户的不感兴趣标签、以及表征该每个用户的浏览行为的数据。根据本发明的另一些实施例,该表征每个用户的浏览行为的数据,例如可以是用户的浏览行为所涉及的信息所包含标签信息、每个标签对应的点击次数、投诉次数、忽略次数、好评次数等,和/或用户浏览每个信息时的地理位置信息、时间信息等。

[0069] 具体地,以表1的原始数据为例。该第一类特征数据可以从表1所示的原始数据中提取出“用户ID”、“年龄”、“兴趣标签”、“文章ID”以及对应的“浏览行为”的数据,并进行结构化和规范化处理得到。该第一类特征数据中的特征,旨在描述用户对信息进行点击的原因。

[0070] 该第二类特征数据可以从表1所示的原始数据中提取出“用户ID”、“年龄”、“不感兴趣标签”、“文章ID”以及对应的“浏览行为”数据,并进行结构化和规范化处理得到。该第二类特征数据中的特征,旨在描述用户对信息不进行点击的原因。

[0071] 接着,在操作S303,通过机器学习获得该第一模型和该第二模型。其中,跟本发明的实施例,操作S303的具体实现可以参考图4的示意。

[0072] 图4示意性地示出了根据本发明一实施例的操作S303中通过机器学习获得第一模型和第二模型的流程图。

[0073] 如图4所示,根据本发明的实施例,操作S303可以包括操作S413和操作S423。

[0074] 其中,在操作S413,通过第一类特征数据训练该第一模型。以及,在操作S423,通过第二类特征数据训练该第二模型。训练后得到的该第一模型可以用于预测用户点击信息的行为,第二模型可以用于预测用户不点击信息的行为。此后,就可以使用该第一模型和第二模型进行用户点击行为或不点击行为的预测。

[0075] 以此方式,根据本发明的实施例,通过第一模型和第二模型分别预测第一用户点击该第一信息的点击概率和第一用户不点击该第一信息的不点击概率,然后在该点击概率大于不点击概率时,认为用户会点击该第一信息,从而将该第一信息作为推荐的备选信息,有效提高信息推荐的针对性和有效性。

[0076] 本发明对第一模型和第二模型不进行限制,任何能预测第一用户点击该第一信息的点击概率和第一用户不点击该第一信息的不点击概率的模型都适用于本发明。例如,第一模型和第二模型可以为logistics回归模型,但是采用的特征数据不一样。

[0077] 图5示意性地示出了根据本发明实施例的用于信息推荐方法的构思示意图。

[0078] 如图5所示,根据本发明实施例的用于信息推荐的方法,旨在对于同一信息,预测得到用户点击该信息的点击概率和用户不点击该信息的不点击概率,然后根据该点击概率和不点击概率的比较来决定是否保留该信息以进行推荐,避免现有技术中通过经验决策导致信息推荐结果的不可控性。

[0079] 假设数据库中有大量信息,针对不同的用户要确定向每个用户推荐那些信息,不推荐哪些信息。以第一用户来指代任意的一个用户,以第一信息来指代该大量信息中的任意一个信息。

[0080] 参考图5的示意,该第一模型的学习目标为对每个用户的点击行为进行建模,从而可以用来预测每个用户(例如,第一用户)点击第一信息的点击概率。

[0081] 该第一模型主要构建流程为,首先读取原始数据,该原始数据来自大量用户的历史浏览行为数据。然后,提取第一类特征数据。第一类特征数据的特征旨在描述该网络平台中每个用户对信息进行点击的原因,其中,该第一类特征数据的具体内容如前所述。最终,通过第一类特征数据训练第一模型,从而可以使第一模型学习到每个用户的点击行为。

[0082] 然后在预测阶段,针对第一用户,通过第一模型可以预测获得该第一用户点击第一信息的点击概率。

[0083] 继续参考图5,该第二模型的学习目标为对每个用户的不点击行为进行建模,从而可以用来预测第一用户不点击第一信息的不点击概率。

[0084] 该第二模型的主要构建流程为,首先读取该原始数据。然后,从原始数据总提取第二类特征数据。该第二类特征数据的特征旨在描述该网络平台中每个用户对信息不点击的原因,其中,该第二类特征数据的具体内容如前所述。最后,通过第二类特征数据训练第二模型,从而可以使第二模型学习到每个用户的不点击行为。

[0085] 然后在预测阶段,针对第一用户,通过第二模型可以预测获得该第一用户不点击第一信息的不点击概率。

[0086] 继续参考图5,最后,将该第一用户对第一信息的点击概率与不点击概率进行比较,当点击概率大于不点击概率时将该第一信息作为推荐备选信息。此处的理论依据为,当第一用户对第一信息进行点击的概率大于其不进行点击的概率时,可以认为用户才会对一个样本真正进行点击。

[0087] 以此类推,对于数据库中大量待筛选的信息中的每一个,都按照图2~图5示意的的方法获得点击概率和不点击概率,并在点击概率大于不点击概率时保留该信息,从而可以有效地降低使用先验知识进行筛选带来的不可控性,提高信息推荐的针对性和有效性。

[0088] 示例性装置

[0089] 在介绍了本发明示例性实施方式的方法之后,接下来,参考图6对本发明示例性实施方式的用于信息推荐的装置进行介绍。

[0090] 图6示意性地示出了根据本发明实施例的用于信息推荐的装置600的框图。

[0091] 如图6所示,该用于信息推荐的装置600包括点击概率预测模块610、不点击概率预

测模块620以及决策模块630。

[0092] 点击概率预测模块610用于通过第一模型预测第一用户点击第一信息的概率,得到点击概率(操作S201)。不点击概率预测模块620用于通过第二模型预测该第一用户不点击该第一信息的概率,得到不点击概率,其中,该第一模型与第二模型不同(操作S202)。决策模块用于在该点击概率大于该不点击概率时,将该第一信息作为向该第一用户推荐的备选信息(操作S203)。

[0093] 根据本发明的一个实施例,该第一模型为用于预测用户点击信息的行为的模型,以及该第二模型为用于预测用户不点击信息的行为的模型。

[0094] 根据本发明的另一个实施例,该装置600还包括模型获得模块640。模型获得模块640用于通过机器学习获得该第一模型和该第二模型(操作S303)。

[0095] 该模型获取模块640具体可以包括第一训练子模块641以及第二训练子模块642。第一训练子模块641用于通过第一类特征数据训练该第一模型(操作S413)。第二训练子模块642用于通过第二类特征数据训练该第二模型(操作S423)。其中,该第一类特征数据包括每个用户的兴趣标签、以及表征该每个用户的浏览行为的数据,该第二类特征数据包括每个用户的不感兴趣标签、以及该表征该每个用户的浏览行为的数据。根据本发明的一个实施例,该浏览行为包括以下任意一种或多种行为:点击信息的行为、评价信息的行为、浏览标题但未点击信息的行为、取消关注信息的行为、以及强制关闭信息的行为。

[0096] 根据本发明的又一个实施例,该装置600还包括原始数据获取模块650和特征数据提取模块660。原始数据获取模块650用于获取原始数据,该原始数据包括特定时间内预定网络平台中多个用户的与该浏览行为相关的数据(操作S301)。特征数据提取模块660用于从该原始数据中提取该第一类特征数据和该第二类特征数据(操作S302)。

[0097] 根据本发明的实施,该用于信息推荐的装置600可以用于实现如上所述根据本发明实施例的用于信息推荐的方法,提高信息推荐的针对性和有效性,详细内容可以参考图2~图5的描述,在此不再赘述。

#### [0098] 示例性介质

[0099] 在介绍了本发明示例性实施方式的方法和装置之后,接下来,参考图7对本发明示例性实施方式的用于信息推荐的程序产品(亦可以称之为计算机可读存储介质)进行说明。

[0100] 根据本发明的实施例,还提供了一种计算机可读存储介质,其上存储有可执行指令,所述指令在被处理器执行时使处理器执行根据本发明实施例的用于信息推荐的方法。

[0101] 在一些可能的实施方式中,本发明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,当所述程序产品在计算设备上运行时,所述程序代码用于使所述计算设备执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的用于信息推荐的方法中的步骤,例如,所述计算设备可以执行如图2中所示的操作S201:通过第一模型预测第一用户点击第一信息的概率,得到点击概率;操作S202:通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率;以及操作S203:在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息;其中,所述第一模型与所述第二模型不同。

[0102] 所述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。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。

[0103] 如图7所示,描述了根据本发明的实施方式的用于信息推荐的程序产品700,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,并可以在计算设备,例如个人电脑上运行。然而,本发明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。

[0104] 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发明操作的程序代码,所述程序设计语言包括面向对象的程序设计语言——诸如Java,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] 示例性计算设备

[0108] 在介绍了本发明示例性实施方式的方法、装置和介质之后,接下来,参考图8对本发明示例性实施方式的用于信息推荐的计算设备进行说明。

[0109] 根据本发明的实施例,还提供了一种计算设备。所述计算设备包括存储有可执行指令的一个或多个存储器,以及一个或多个处理器。所述处理器执行所述可执行指令,用以实现根据本发明实施例的用于信息推荐的方法。

[0110] 本发明实施例还提供了一种计算设备。所属技术领域的技术人员能够理解,本发明的各个方面可以实现为系统、方法或程序产品。因此,本发明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”。

[0111] 在一些可能的实施方式中,根据本发明的计算设备可以至少包括至少一个处理单元、以及至少一个存储单元。其中,所述存储单元存储有程序代码,当所述程序代码被所述处理单元执行时,使得所述处理单元执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的用于信息推荐的操作。例如,所述处理单元可以执行如图2中所示的操作S201:通过第一模型预测第一用户点击第一信息的概率,得到点击概率;操作S202:通过第二模型预测所述第一用户不点击所述第一信息的概率,得到不点击概率;以及

操作S203:在所述点击概率大于所述不点击概率时,将所述第一信息作为向所述第一用户推荐的备选信息;其中,所述第一模型与所述第二模型不同。

[0112] 下面参照图8来描述根据本发明的这种实施方式的用于信息推荐的计算设备800。如图8所示的计算设备800仅仅是一个示例,不对本发明实施例的功能和使用范围带来任何限制。

[0113] 如图8所示,计算设备800以通用计算设备的形式表现。计算设备800的组件可以包括但不限于:上述至少一个处理单元810、上述至少一个存储单元820、连接不同系统组件(包括存储单元820和处理单元810)的总线830。

[0114] 总线830包括数据总线、控制总线和地址总线。

[0115] 存储单元820可以包括易失性存储器,例如随机存取存储器(RAM)821和/或高速缓存存储器822,还可以进一步包括只读存储器(ROM)823。

[0116] 存储单元820还可以包括具有一组(至少一个)程序模块824的程序/实用工具825,这样的程序模块824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。

[0117] 计算设备800也可以与一个或多个外部设备840(例如键盘、指向设备、蓝牙设备等)通信,这种通信可以通过输入/输出(I/O)接口850进行。并且,计算设备800还可以通过网络适配器860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860通过总线830与计算设备800的其它模块通信。应当明白,尽管图中未示出,可以结合计算设备800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。

[0118] 应当注意,尽管在上文详细描述中提及了装置的若干单元/模块或子单元/模块,但是这种划分仅仅是示例性的并非强制性的。实际上,根据本发明的实施方式,上文描述的两个或更多单元/模块的特征和功能可以在一个单元/模块中具体化。反之,上文描述的一个单元/模块的特征和功能可以进一步划分为由多个单元/模块来具体化。

[0119] 此外,尽管在附图中以特定顺序描述了本发明方法的操作,但是,这并非要求或者暗示必须按照该特定顺序来执行这些操作,或是必须执行全部所示的操作才能实现期望的结果。附加地或备选地,可以省略某些步骤,将多个步骤合并为一个步骤执行,和/或将一个步骤分解为多个步骤执行。

[0120] 虽然已经参考若干具体实施方式描述了本发明的精神和原理,但是应该理解,本发明并不限于所公开的具体实施方式,对各方面的划分也不意味着这些方面中的特征不能组合以进行受益,这种划分仅是为了表述的方便。本发明旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。

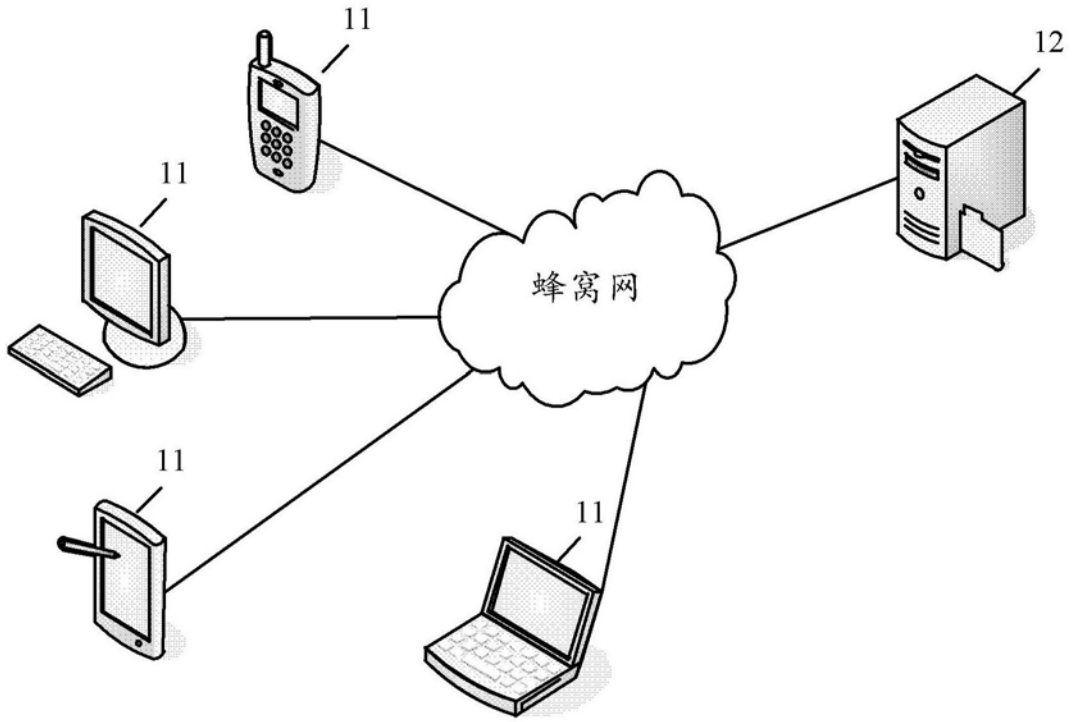


图1

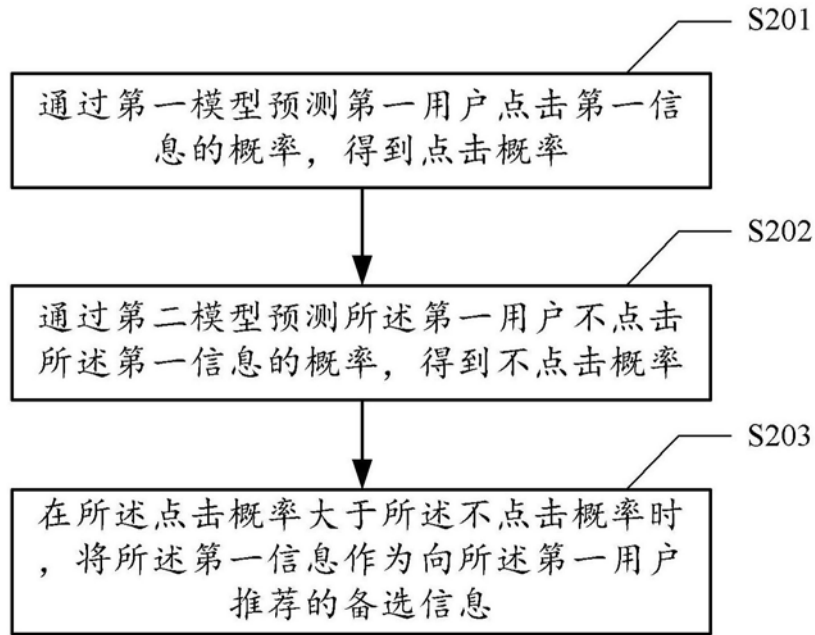


图2

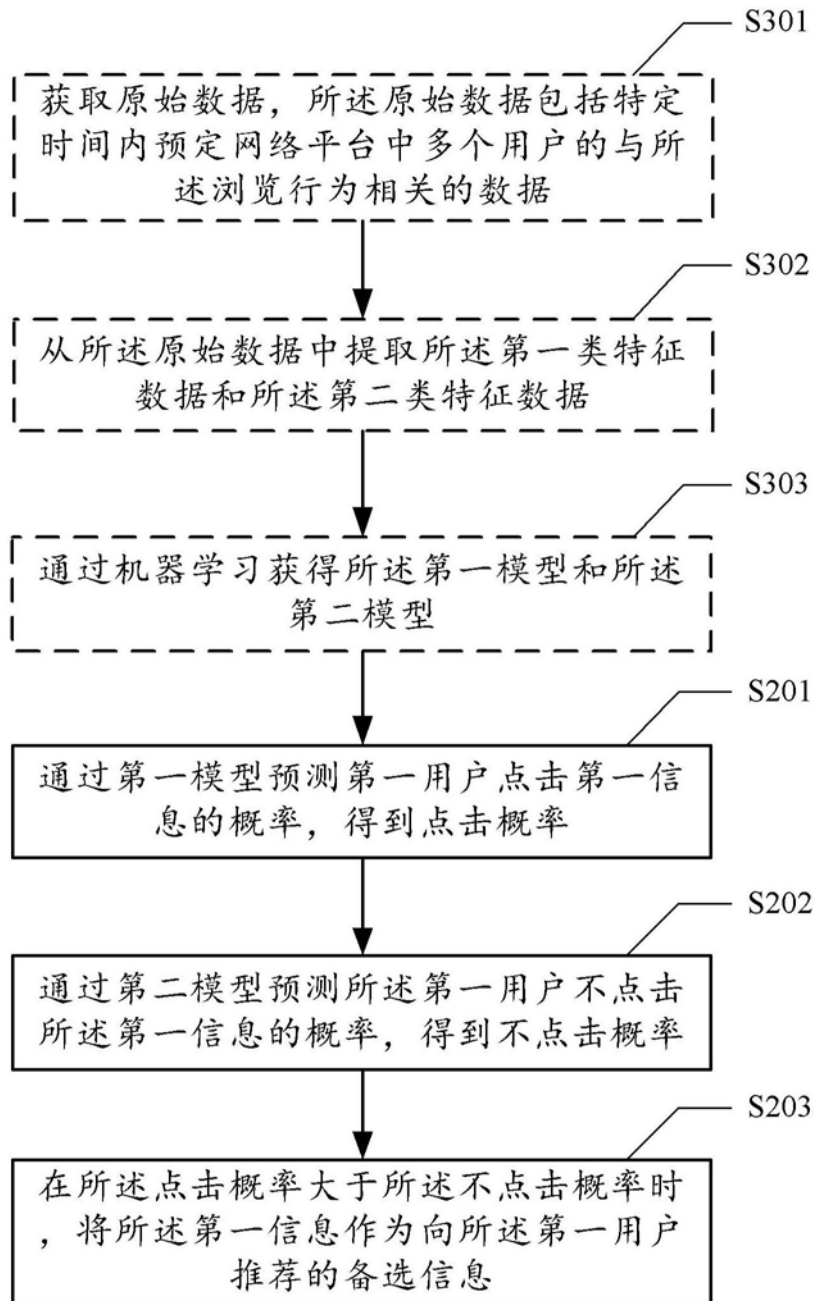


图3



S303

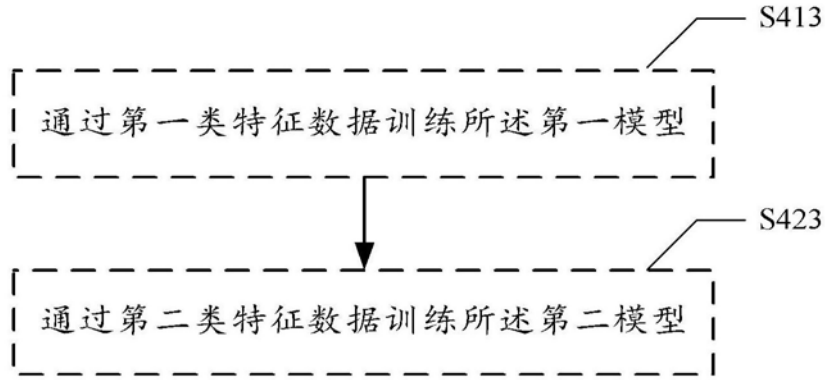


图4

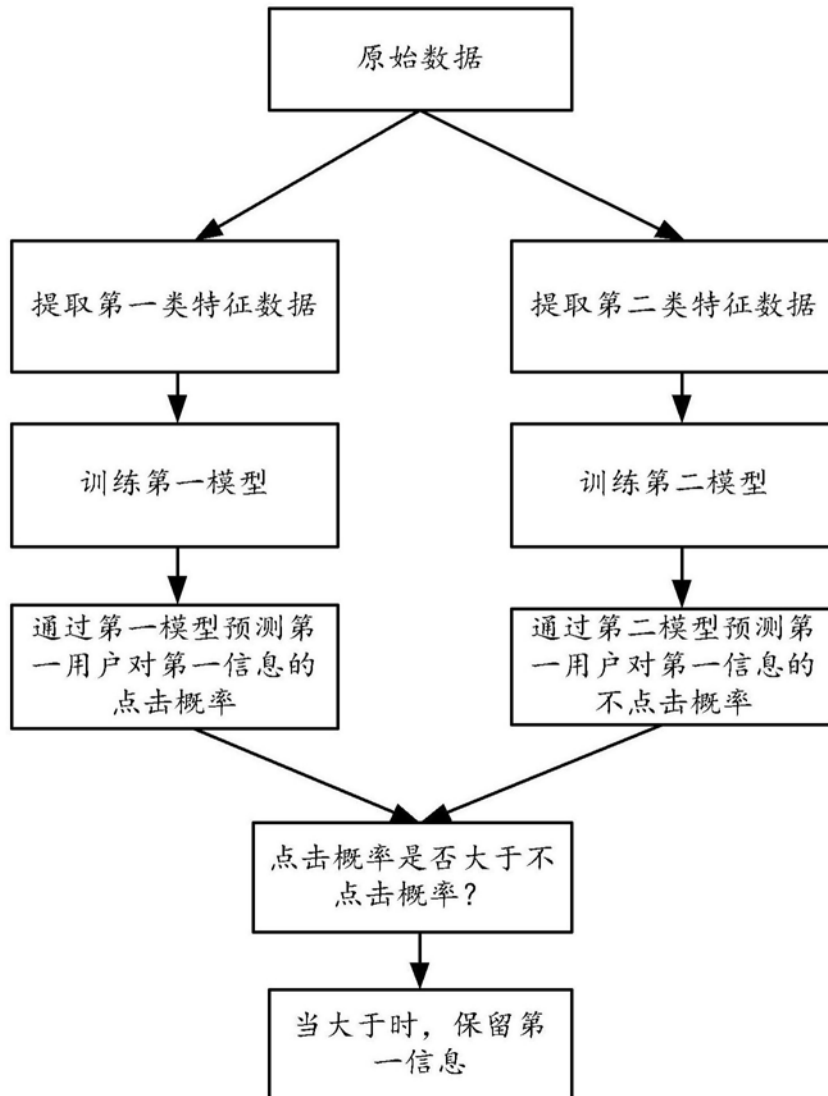


图5

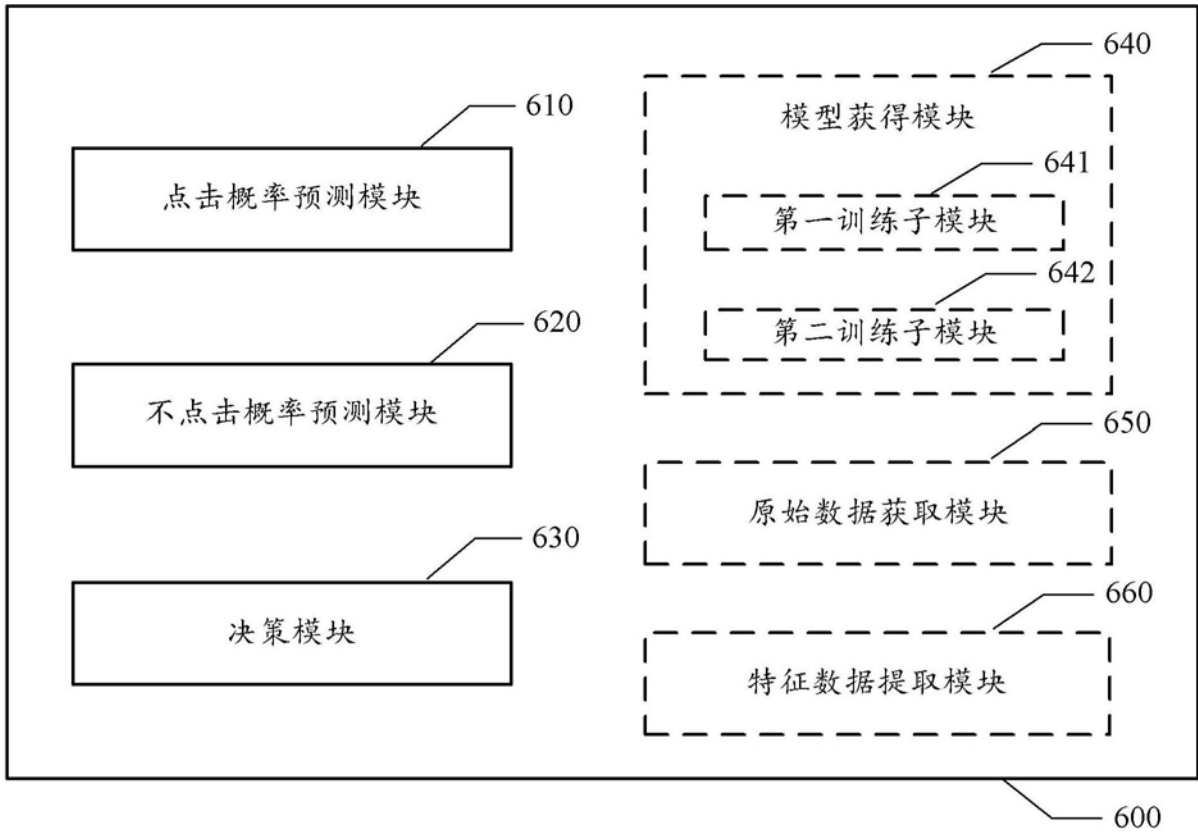


图6

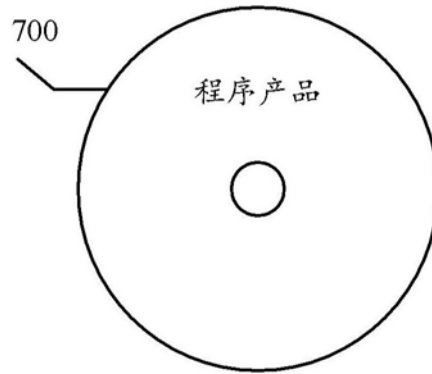


图7

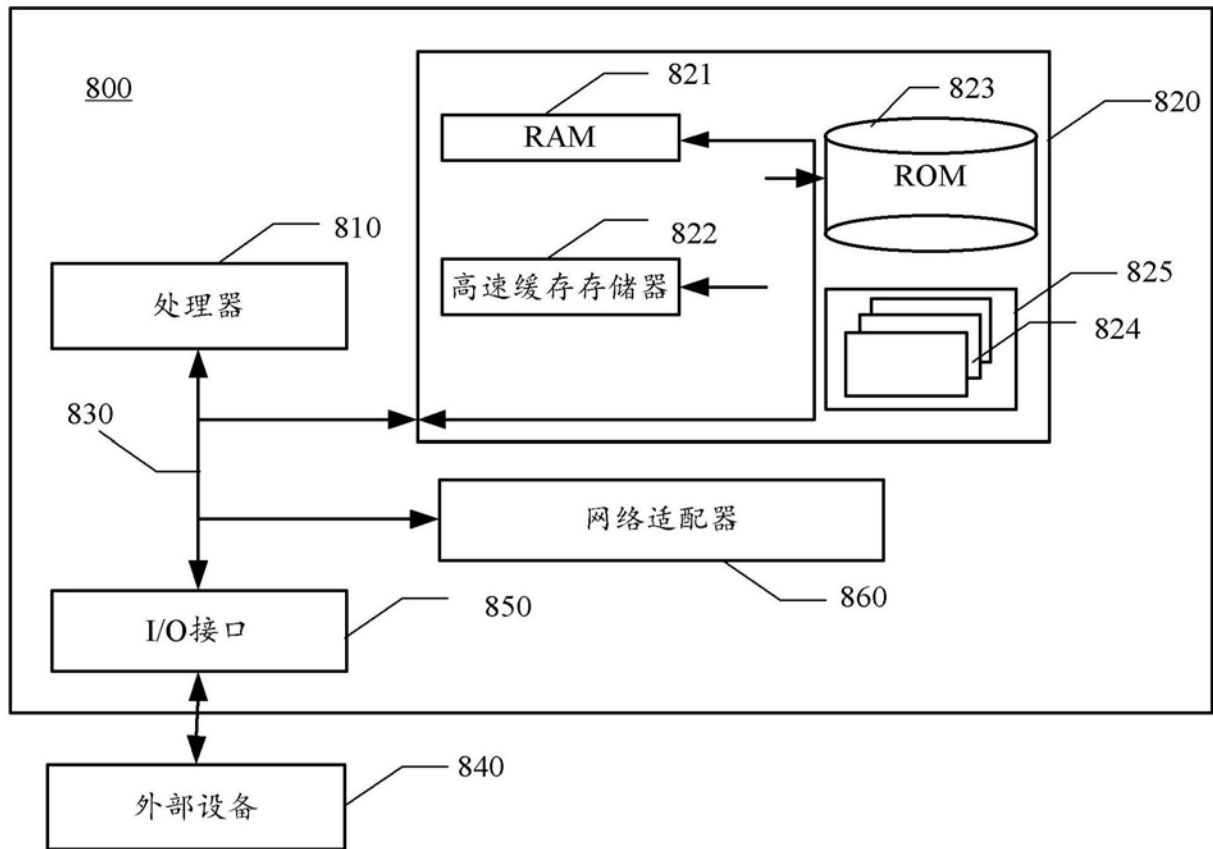


图8