



(12)发明专利

(10)授权公告号 CN 106503014 B

(45)授权公告日 2020.08.07

(21)申请号 201510564481.3

(22)申请日 2015.09.08

(65)同一申请的已公布的文献号
申请公布号 CN 106503014 A

(43)申请公布日 2017.03.15

(73)专利权人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市福田区振兴路
赛格科技园2栋东403室

(72)发明人 胡雨成

(74)专利代理机构 深圳翼盛智成知识产权事务
所(普通合伙) 44300
代理人 黄威

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

(56)对比文件

US 8001105 B2,2011.08.16

US 2014324579 A1,2014.10.30

审查员 汤明达

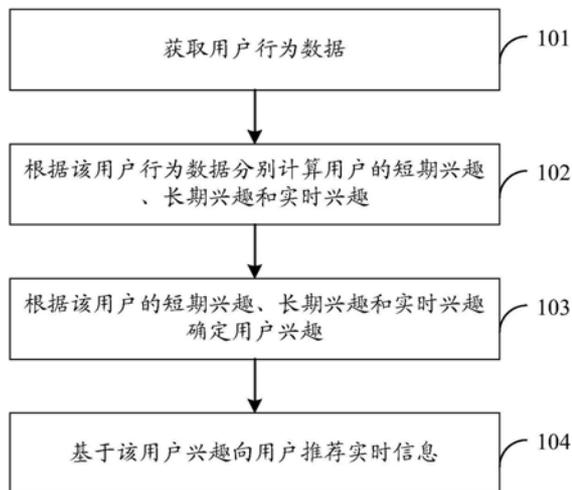
权利要求书4页 说明书12页 附图4页

(54)发明名称

一种实时信息的推荐方法、装置和系统

(57)摘要

本发明实施例公开了一种实时信息的推荐方法、装置和系统;本发明实施例采用获取用户行为数据,根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,然后根据用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,并基于该用户兴趣向用户推荐实时信息;该方案可以提高时效性,以及改善推荐效果。



1. 一种实时信息的推荐方法,其特征在于,包括:
 - 获取用户行为数据;
 - 根据所述用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣;
 - 根据所述用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣;
 - 从原始实时信息库中获取原始新闻,对获取到的原始新闻进行特征提取,根据提取到的特征对该原始新闻进行分类预测,确定原始新闻的类目;
 - 采用LDA主题模型识别所述原始新闻的文档中潜藏的主题信息,以确定所述原始新闻的主题;
 - 对获取到的原始新闻的内容进行词性加权处理后,进行文本域加权,以确定原始新闻所属的关键字;
 - 根据所述原始新闻的类目、主题和关键字计算该原始实时信息库中原始新闻的倒排索引,得到新闻的倒排索引;
 - 根据所述用户兴趣,从新闻的倒排索引中召回相应的原始新闻,得到候选推荐信息;
 - 基于所述候选推荐信息推荐实时新闻给所述用户。
2. 根据权利要求1所述的方法,其特征在于,所述根据所述用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,包括:
 - 根据所述用户行为数据计算用户在预置周期中每一天的兴趣权值,得到天兴趣权值,并对所述天兴趣权值按照时间进行衰减,得到用户的短期兴趣;
 - 根据所述用户行为数据计算用户在预置时间范围内的兴趣权值,得到用户的长期兴趣,所述预置时间范围大于一天;
 - 根据所述用户行为数据确定用户当前点击的兴趣权值,得到用户的实时兴趣。
3. 根据权利要求2所述的方法,其特征在于,所述对所述天兴趣权值按照时间进行衰减,得到用户的短期兴趣,包括:
 - 根据所述天兴趣权值确定当前需要进行衰减的兴趣权值;
 - 对所述需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值;
 - 返回执行根据所述天兴趣权值确定当前需要进行衰减的兴趣权值的步骤,直至所述天兴趣权值中所有需要进行衰减的兴趣权重衰减完毕;
 - 对得到的所有衰减后的兴趣权值进行统计,得到用户的短期兴趣。
4. 根据权利要求3所述的方法,其特征在于,所述对所述需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值,包括:
 - 确定所述需要进行衰减的兴趣权值所在的日期与当前日期的日期差;
 - 计算所述日期差与预置衰减系数的乘积,并计算1与所述乘积的差;
 - 将所述需要进行衰减的兴趣权值乘以所述差,得到衰减后的兴趣权值。
5. 根据权利要求2所述的方法,其特征在于,所述根据所述用户行为数据计算用户在预置时间范围内的兴趣权值,得到用户的长期兴趣,包括:
 - 根据所述用户行为数据统计用户在当前日期的一年内,每个月的用户行为;
 - 根据所述每个月的用户行为计算每个兴趣在当月的权重;
 - 根据所述每个兴趣在当月的权重计算一年内每个兴趣的平均权重;
 - 对所述平均权重进行统计,得到用户的长期兴趣。

6. 根据权利要求1所述的方法,其特征在于,所述将基于所述候选推荐信息推荐实时新闻给所述用户,包括:

计算所述候选推荐信息中各实时新闻与所述用户兴趣的匹配度,得到实时新闻的兴趣相关性;

确定所述候选推荐信息中各实时新闻的发布时间确定各实时新闻的新鲜度,得到实时新闻的时新性;

确定所述候选推荐信息中各实时新闻的点击率,并根据所述点击率计算点击模型因子;

根据所述兴趣相关性、时新性和点击模型因子从所述候选推荐信息中确定推荐信息;将所述推荐信息推荐给所述用户。

7. 根据权利要求6所述的方法,其特征在于,所述根据所述兴趣相关性、时新性和点击模型因子从所述候选推荐信息中确定推荐信息,包括:

根据所述兴趣相关性、时新性和点击模型因子对所述候选推荐信息中的实时信息进行评分;

将评分高于预置阈值的实时信息确定为推荐信息。

8. 根据权利要求6所述的方法,其特征在于,所述根据所述兴趣相关性、时新性和点击模型因子从所述候选推荐信息中确定推荐信息之前,还包括:

确实所述候选推荐信息中各实时新闻的信息质量;

所述根据所述兴趣相关性、时新性和点击模型因子从所述候选推荐信息中确定推荐信息,包括:根据所述兴趣相关性、时新性、点击模型因子和信息质量从所述候选推荐信息中确定推荐信息。

9. 根据权利要求8所述的方法,其特征在于,所述根据所述兴趣相关性、时新性、点击模型因子和信息质量从所述候选推荐信息中确定推荐信息,包括:

根据所述兴趣相关性、时新性、点击模型因子和信息质量对所述候选推荐信息中的实时新闻进行评分;

将评分高于预置阈值的实时新闻确定为推荐信息。

10. 根据权利要求1所述的方法,其特征在于,所述根据所述用户兴趣,从新闻的倒排索引中召回相应的原始新闻,得到候选推荐信息,具体为:根据用户兴趣确定用户所感兴趣的类目、主题和/或关键字,从所述新闻的倒排索引中获取与用户所感兴趣的类目、主题和/或关键字相同、相似或相近的原始新闻,得到候选推荐信息。

11. 一种实时信息的推荐装置,其特征在于,包括:

获取单元,用于获取用户行为数据;

运算单元,用于根据所述用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣;

确定单元,用于根据所述用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣;

提取单元,用于从原始实时信息库中获取原始新闻,对获取到的原始新闻进行特征提取,根据提取到的特征对该原始新闻进行分类预测,确定原始新闻的类目;

识别单元,用于采用LDA主题模型识别所述原始新闻的文档中潜藏的主题信息,以确定所述原始新闻的主题;

对获取到的原始新闻的内容进行词性加权处理后,进行文本域加权,以确定原始新闻所属的关键字;

倒排索引确定单元,用于根据所述原始新闻的类目、主题和关键字计算该原始实时信息库中原始新闻的倒排索引,得到新闻的倒排索引;根据所述用户兴趣,从新闻的倒排索引中召回相应的原始新闻,得到候选推荐信息;

推荐单元,用于基于所述候选推荐信息推荐实时新闻给所述用户。

12. 根据权利要求11所述的装置,其特征在于,所述运算单元包括第一计算子单元、第二计算子单元和第三计算子单元;

第一计算子单元,用于根据所述用户行为数据计算用户在预置周期中每一天的兴趣权值,得到天兴趣权值,并对所述天兴趣权值按照时间进行衰减,得到用户的短期兴趣;

第二计算子单元,用于根据所述用户行为数据计算用户在预置时间范围内的兴趣权值,得到用户的长期兴趣,所述预置时间范围大于一天;

第三计算子单元,用于根据所述用户行为数据确定用户当前点击的兴趣权值,得到用户的实时兴趣。

13. 根据权利要求12所述的装置,其特征在于,

所述第一计算子单元,具体用于根据所述天兴趣权值确定当前需要进行衰减的兴趣权值,对所述需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值;返回执行根据所述天兴趣权值确定当前需要进行衰减的兴趣权值的操作,直至所述天兴趣权值中所有需要进行衰减的兴趣权重衰减完毕;对得到的所有衰减后的兴趣权值进行统计,得到用户的短期兴趣。

14. 根据权利要求13所述的装置,其特征在于,

所述第一计算子单元,具体用于确定所述需要进行衰减的兴趣权值所在的日期与当前日期的日期差;计算所述日期差与预置衰减系数的乘积,并计算1与所述乘积的差;将所述需要进行衰减的兴趣权值乘以所述1与所述乘积的差,得到衰减后的兴趣权值。

15. 根据权利要求13所述的装置,其特征在于,

第二计算子单元,具体用于根据所述用户行为数据统计用户在当前日期的一年内,每个月的用户行为;根据所述每个月的用户行为计算每个兴趣在当月的权重;根据所述每个兴趣在当月的权重计算一年内每个兴趣的平均权重;对所述平均权重进行统计,得到用户的长期兴趣。

16. 根据权利要求11所述的装置,其特征在于,所述推荐单元包括推荐子单元,

所述推荐子单元,具体用于计算所述候选推荐信息中各实时新闻与所述用户兴趣的匹配度,得到实时新闻的兴趣相关性;确定所述候选推荐信息中各实时新闻的发布时间确定各实时新闻的新鲜度,得到实时新闻的时新性;确定所述候选推荐信息中各实时新闻的点击率,并根据所述点击率计算点击模型因子;根据所述兴趣相关性、时新性和点击模型因子从所述候选推荐信息中确定推荐信息;将所述推荐信息推荐给所述用户。

17. 根据权利要求11所述的装置,其特征在于,所述推荐单元包括推荐子单元,

所述推荐子单元,具体用于计算所述候选推荐信息中各实时新闻与所述用户兴趣的匹配度,得到实时新闻的兴趣相关性;确定所述候选推荐信息中各实时新闻的发布时间确定各实时信息的新鲜度,得到实时新闻的时新性;确定所述候选推荐信息中各实时新闻的点

击率,并根据所述点击率计算点击模型因子;确实所述候选推荐信息中各实时新闻的信息质量;根据所述兴趣相关性、时新性、点击模型因子和信息质量从所述候选推荐信息中确定推荐信息;将所述推荐信息推荐给所述用户。

18. 根据权利要求11所述的装置,其特征在于,所述推荐单元包括召回子单元,所述召回子单元,具体用于根据用户兴趣确定用户所感兴趣的类目、主题和/或关键字,从所述实时信息的倒排索引中获取与用户所感兴趣的类目、主题和/或关键字相同、相似或相近的原始实时信息,得到候选推荐信息。

19. 一种实时信息的推荐系统,其特征在于,包括权利要求11至18所述的任一种实时信息的推荐装置。

20. 一种计算机存储介质,其特征在于,所述计算机存储介质存储有多条指令,所述指令适于处理器进行加载,以执行权利要求1至10任一项所述的实时信息的推荐方法。

一种实时信息的推荐方法、装置和系统

技术领域

[0001] 本发明涉及通信技术领域,具体涉及一种实时信息的推荐方法、装置和系统。

背景技术

[0002] 互联网信息的急速增长,使得人们淹没在资讯的海洋中,如何在浩瀚的信息海洋中快速地找到需要的信息,是一个极为重要的问题。

[0003] 为了解决互联网海量信息资源出现的“信息过载”问题,协助用户从琳琅满目的海量信息中快速获取到自己所需的信息,现有技术提出了各种信息的推荐方案,所谓信息推荐,指的是根据用户的兴趣特点和行为,向用户推荐用户感兴趣的信息,为此,现有的推荐算法主要可分为两大类,一是基于行为的推荐算法、二是基于内容的推荐算法。其中,基于行为的推荐算法主要是通过统计用户对信息的行为,以及计算推荐池中各信息的相似度,然后将与该行为所对应的信息相似度较高的信息推荐给用户。而基于内容的推荐算法则主要是通过对各类信息进行打关键字,以及对用户的兴趣进行挖掘,以确定用户所感兴趣的关键字,然后基于该感兴趣的关键字和各类信息的关键字计算推荐列表,并推荐给用户。

[0004] 在对现有技术的研究和实践过程中,本发明的发明人发现,现有的推荐方案或是需要依赖大量的用户参与,或时没有考虑用户的兴趣变化,因此,时效性较差,而对于实时信息,如新闻类的信息,均具有一次性消费的特点(即对于相同内容的新闻,用户只会阅读一次),时效性极为重要,因此,对于实时信息而言,现有的推荐方案的推荐效果并不佳。

发明内容

[0005] 本发明实施例提供一种实时信息的推荐方法、装置和系统,可以提高其时效性,灵活、准确且及时地将用户当前最感兴趣的实时信息推荐给用户,大大改善推荐效果。

[0006] 本发明实施例提供一种实时信息的推荐方法,包括:

[0007] 获取用户行为数据;

[0008] 根据所述用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣;

[0009] 根据所述用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣;

[0010] 基于所述用户兴趣向用户推荐实时信息。

[0011] 相应的,本发明实施例还提供一种实时信息的推荐装置,包括:

[0012] 获取单元,用于获取用户行为数据;

[0013] 运算单元,用于根据所述用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣;

[0014] 确定单元,用于根据所述用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣;

[0015] 推荐单元,用于基于所述用户兴趣向用户推荐实时信息。

[0016] 此外,本发明实施例还提供一种实时信息的推荐系统,包括本发明实施例提供的任一种实时信息的推荐装置。

[0017] 本发明实施例采用获取用户行为数据,根据该用户行为数据分别计算用户的短期

兴趣、长期兴趣和实时兴趣,然后根据用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,并基于该用户兴趣向用户推荐实时信息;由于该方案在计算用户的兴趣时,不仅将用户的长期兴趣作为考量因素,而且,也将用户的短期兴趣与实时兴趣也作为考量因素,以反映出用户的兴趣随时间变化的情况,因此,相对于现在技术而言,可以更加灵活、准确且及时地将用户当前最感兴趣的实时信息推荐给用户,在提高时效性的同时,大大改善推荐效果。

附图说明

[0018] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域技术人员来讲,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。

[0019] 图1a是本发明实施例提供的实时信息的推荐系统的场景示意图;

[0020] 图1b是本发明实施例提供的实时信息的推荐方法的流程图;

[0021] 图2是本发明实施例提供的实时信息的推荐方法的另一流程图;

[0022] 图3a是本发明实施例提供的实时信息的推荐装置的结构图;

[0023] 图3b是本发明实施例提供的实时信息的推荐装置的另一结构图。

具体实施方式

[0024] 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,所描述的实施例仅仅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。

[0025] 本发明实施例提供一种实时信息的推荐方法、装置和系统。

[0026] 参见图1a,该实时信息的推荐系统包括本发明实施例所提供的任一种实时信息的推荐装置,该实时信息的推荐装置具体可以集成在服务器,比如推荐服务器中。此外,该实时信息的推荐系统还可以包括其他的设备,比如用户设备,以及用于保存用户行为数据的用户服务器,以及用于保存原始实时信息的信息服务器等。

[0027] 例如,以该实时信息的推荐装置集成在推荐服务器中为例,则当需要对实时信息进行推荐时,该推荐服务器可以从用户服务器获取用户行为数据,根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,然后根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,从信息服务器获取实时信息,并基于该用户兴趣向用户设备推荐实时信息,比如新闻等。

[0028] 其中,短期兴趣指的是用户在较短的一个周期内所计算得到的兴趣权值,具体可以根据获取到的用户行为数据计算用户在预置周期(比如30天)中每一天的兴趣权值,得到天兴趣权值,并对该天兴趣权值按照时间进行衰减来得到;长期兴趣指的是用户在较长一个周期内所计算得到的兴趣权重,比如可以根据该用户行为数据计算用户在一年内的兴趣权值,等等;而实时兴趣则指的是用户当前兴趣权值,比如用户当前点击的了某个关键字或标签,等等。

[0029] 以下将分别进行详细说明。

[0030] 实施例一、

[0031] 本实施例将从实时信息的推荐装置的角度进行描述,该实时信息的推荐装置具体可以集成在服务器,比如推荐服务器等设备中。

[0032] 一种实时信息的推荐方法,包括:获取用户行为数据;根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣;根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣;基于该用户兴趣向用户推荐实时信息。

[0033] 如图1b所示,该实时信息的推荐方法的具体流程可以如下:

[0034] 101、获取用户行为数据。

[0035] 其中,用户行为数据指的是可供用户行为分析的相关数据,比如用户的浏览记录、点击记录和/或下载记录等数据。这些用户行为数据可以存储在该实时信息的推荐装置中,也可以存储在其他设备,比如用户服务器等设备中。

[0036] 以存储在用户服务器为例,则此时,具体可以从用户服务器中获取用户行为数据。

[0037] 102、根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣。例如,具体可以如下:

[0038] (1) 根据该用户行为数据计算用户在预置周期中每一天的兴趣权值,得到天兴趣权值,并对该天兴趣权值按照时间进行衰减,得到用户的短期兴趣。

[0039] 其中,按时间进行衰减的方式可以有多种,例如,具体可以如下:

[0040] A、根据该天兴趣权值确定当前需要进行衰减的兴趣权值。

[0041] B、对该需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值。

[0042] 比如,可以确定该需要进行衰减的兴趣权值所在的日期与当前日期的日期差,计算该日期差与预置衰减系数的乘积,并计算1与该乘积的差,然后将该需要进行衰减的兴趣权值乘以该差,得到衰减后的兴趣权值。

[0043] 比如,若用 μ 来表示衰减系数,用D表示需要进行衰减的兴趣权值所在的日期与当前日期的日期差, $interest_weight_k$ 表示需要进行衰减的兴趣权值,则衰减后兴趣权值 $final_weight_k$ 用公式表示为:

[0044] $final_weight_k = interest_weight_k * (1 - D * \mu)$ 。

[0045] 其中,衰减系数可以根据实际应用的需求进行设置,在此不再赘述。

[0046] C、返回执行根据该天兴趣权值确定当前需要进行衰减的兴趣权值的步骤,直至该天兴趣权值中所有需要进行衰减的兴趣权重衰减完毕。

[0047] 即对每一天的兴趣权值均按照时间进行衰减处理,得到对每一天应的衰减后的兴趣权值。

[0048] D、对得到的所有衰减后的兴趣权值进行统计,得到用户的短期兴趣。

[0049] 其中,预置周期可以根据实际应用的需求进行设置,比如,一般可以设置为7天、15天或30天等。

[0050] 例如,以30天为例,则若当前日期为8月10号,则可以以8月10号为基准,获取30天,即7月12号至8月10号这段日期中每一天的兴趣权值,然后,对该每一天的兴趣权值按照时间进行衰减,比如,以8月8日为例,由于8月8日与8月10号的日期差为两天,则8月8日这一天的衰减后兴趣权值 $final_weight_{28}$ 为:

[0051] $final_weight_{28} = interest_weight_{28} * (1 - 2 * \mu)$ 。

[0052] 当然,也可以采用其他的衰减方式,在此不再赘述。

[0053] (2) 根据该用户行为数据计算用户在预置时间范围内的兴趣权值,得到用户的长期兴趣。

[0054] 其中,该预置时间范围可以根据实际应用的需求进行设置,至少大于一天,比如,可以设置为一个季度,一年或两年,等等。

[0055] 例如,以一年为例,则可以根据该用户行为数据统计用户在当前日期的一年内每个月的用户行为;根据该每个月的用户行为计算每个兴趣在当月的权重;根据该每个兴趣在当月的权重计算一年内每个兴趣的平均权重;对该平均权重进行统计,得到用户的长期兴趣。

[0056] 可选的,也可以对每个兴趣在当月的权重按照时间进行衰减,得到衰减后的兴趣权重,然后对这些衰减后的兴趣权重进行统计,得到用户的长期兴趣,等等。

[0057] (3) 根据该用户行为数据确定用户当前点击的兴趣权值,得到用户的实时兴趣。

[0058] 例如,若用户当前点击的消息包括了“NBA”关键字(或标签),则确定用户当前的兴趣为“NBA”,于是,可以计算该“NBA”关键字的权值,以此类推,便可以得到用户的实时兴趣。

[0059] 103、根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣。

[0060] 例如,可以根据预置的策略将用户的短期兴趣、长期兴趣和实时兴趣进行融合,得到用户兴趣。

[0061] 其中,融合的方式可以有多种,比如,可以分别为短期兴趣、长期兴趣和实时兴趣设置权重,然后根据这些权重将短期兴趣、长期兴趣和实时兴趣融合在一起,又比如,还可以为短期兴趣、长期兴趣和实时兴趣之间的关系设置一个函数,然后通过该函数将短期兴趣、长期兴趣和实时兴趣融合在一起,等等,在此不再赘述。

[0062] 104、基于该用户兴趣向用户推荐实时信息。其中,该实时信息具体可以为新闻等信息。

[0063] 例如,可以根据该用户兴趣,从实时信息的倒排索引中召回相应的实时信息,得到候选推荐信息,基于该候选推荐信息推荐实时信息给该用户,具体可以如下:

[0064] A、计算该候选推荐信息中各实时信息与该用户兴趣的匹配度,得到实时信息的兴趣相关性。

[0065] B、确定该候选推荐信息中各实时信息的发布时间确定各实时信息的新鲜度,得到实时信息的时新性。

[0066] 其中,越新的实时信息,比如越新的新闻,其时新性越高。

[0067] C、确定该候选推荐信息中各实时信息的点击率,并根据该点击率计算点击模型因子(CM)。

[0068] 其中,点击率越大的实时信息,其模型因子越大。

[0069] D、根据该兴趣相关性、时新性和点击模型因子从该候选推荐信息中确定推荐信息。

[0070] 比如,可以根据该兴趣相关性、时新性和点击模型因子对该候选推荐信息中的实时信息进行评分,然后将评分高于预置阈值的实时信息确定为推荐信息。

[0071] 其中,该预置阈值可以根据实际应用的需求进行设置。

[0072] E、将该推荐信息推荐给该用户。

[0073] 可选的,为了提高推荐的信息质量,在确定推荐消息之前,还可以对候选推荐信息中各实时信息的信息质量进行评判,比如,可以通过文本识别的方式确定一个新闻的质量因子,其中,垃圾文章、广告文章的质量分低。即在步骤“根据该兴趣相关性、时新性和点击模型因子从该候选推荐信息中确定推荐信息”之前,该实时信息的推荐方法还可以包括:

[0074] 确实该候选推荐信息中各实时信息的信息质量。

[0075] 则此时,步骤“根据该兴趣相关性、时新性和点击模型因子从该候选推荐信息中确定推荐信息”可以包括:根据上该兴趣相关性、时新性、点击模型因子和信息质量从该候选推荐信息中确定推荐信息,比如,具可以如下:

[0076] 根据该兴趣相关性、时新性、点击模型因子和信息质量对该候选推荐信息中的实时信息进行评分;将评分高于预置阈值的实时信息确定为推荐信息。

[0077] 需说明的是,其中,实时信息的倒排索引可以通过对原始实时信息进行收集和统计来得到,即在步骤“根据该用户兴趣,从实时信息的倒排索引中召回相应的实时信息,得到候选推荐信息”之前,该实时信息的推荐方法还可以包括:

[0078] 从原始实时信息库中获取原始实时信息;对获取到的原始实时信息进行特征提取;根据提取到的特征对该原始实时信息进行分类预测和话题预测,以确定原始实时信息类目和主题;对获取到的原始实时信息的内容进行词性加权处理后,进行文本域加权,以确定原始实时信息所属的关键字;根据原始实时信息类目、主题和关键字计算该原始实时信息库中原始实时信息的倒排索引,得到实时信息的倒排索引。

[0079] 则此时,步骤“根据该用户兴趣,从实时信息的倒排索引中召回相应的实时信息,得到候选推荐信息”具体可以为:根据用户兴趣确定用户所感兴趣的类目、主题和/或关键字,从该实时信息的倒排索引中获取与用户所感兴趣的类目、主题和/或关键字相同、相似或相近的原始实时信息,得到候选推荐信息。

[0080] 例如,可以为用户所感兴趣的类目、主题和/或关键字中所涉及的词语设置同义词和/或近义词,若原始实时信息的类目、主题和/或关键字中包括有与这些同义词和/或近义词相同的词语,则确定该原始实时信息为与用户所感兴趣的类目、主题和/或关键字相似或相近的原始实时信息,等等。

[0081] 由上可知,本实施例采用获取用户行为数据,根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,然后根据用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,并基于该用户兴趣向用户推荐实时信息;由于该方案在计算用户的兴趣时,不仅将用户的长期兴趣作为考量因素,而且,也将用户的短期兴趣与实时兴趣也作为考量因素,以反映出用户的兴趣随时间变化的情况,因此,相对于现在技术而言,可以更加灵活、准确且及时地将用户当前最感兴趣的实时信息推荐给用户,在提高时效性的同时,大大改善推荐效果。

[0082] 此外,在计算候选推荐信息的过程中,不仅可以考虑实时信息的兴趣相关性,同时也可以考虑实时信息的时新性,以及关键字的点击率和信息质量等,因此,可以更加准确的描述这些信息与用户兴趣之间关系强弱,进一步提高了推荐质量和推荐效果。

[0083] 实施例二、

[0084] 根据实施例一所描述的方法,以下将以该实时信息的推荐装置具体集成在推荐服务器,且该实时信息具体为新闻为例进行详细说明。

[0085] 如图2所示,一种实时信息的推荐方法,具体流程可以如下:

[0086] 201、推荐服务器从用户服务器中获取用户行为数据。

[0087] 其中,用户行为数据指的是可供用户行为分析的相关数据,比如用户的浏览记录、点击记录和/或下载记录等数据。

[0088] 202、推荐服务器根据该用户行为数据计算用户在30天内每一天的兴趣权值,得到天兴趣权值,并对该天兴趣权值按照时间进行衰减,得到用户的短期兴趣。例如,具体可以如下:

[0089] A、根据该天兴趣权值确定当前需要进行衰减的兴趣权值。

[0090] B、对该需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值。

[0091] 比如,可以确定该需要进行衰减的兴趣权值所在的日期与当前日期的日期差,计算该日期差与预置衰减系数的乘积,并计算1与该乘积的差,然后将该需要进行衰减的兴趣权值乘以该差,得到衰减后的兴趣权值,用公式表示即为:

[0092] $final_weight_k = interest_weight_k * (1 - D * \mu)$ 。

[0093] 其中, μ 为衰减系数,D为需要进行衰减的兴趣权值所在的日期与当前日期的日期差, $interest_weight_k$ 表示需要进行衰减的兴趣权值, $final_weight_k$ 为衰减后兴趣权值。其中,衰减系数可以根据实际应用的需求进行设置,在此不再赘述。

[0094] C、返回执行根据该天兴趣权值确定当前需要进行衰减的兴趣权值的步骤,直至该天兴趣权值中所有需要进行衰减的兴趣权重衰减完毕。

[0095] 即对每一天的兴趣权值均按照时间进行衰减处理,得到对每一天应的衰减后的兴趣权值。

[0096] D、对得到的所有衰减后的兴趣权值进行统计,得到用户的短期兴趣。

[0097] 例如,若当前日期为8月10号,则可以以8月10为基准,获取30天,即7月12号至8月10号这段日期中每一天的兴趣权值,然后,对该每一天的兴趣权值按照时间进行衰减,比如,以8月8日为例,由于8月8日与8月10号的日期差为两天,则8月8日这一天的衰减后兴趣权值 $final_weight_{28}$ 为:

[0098] $final_weight_{28} = interest_weight_{28} * (1 - 2 * \mu)$ 。

[0099] 当然,也可以采用其他的衰减方式,在此不再赘述。

[0100] 203、推荐服务器根据该用户行为数据统计用户在当前日期的一年内的兴趣权值,得到用户的长期兴趣。

[0101] 例如,具体可以根据该用户行为数据统计用户在当前日期的一年内每个月的用户行为,并根据该每个月的用户行为计算每个兴趣在当月的权重,根据该每个兴趣在当月的权重计算一年内每个兴趣的平均权重,对该平均权重进行统计,得到用户的长期兴趣。

[0102] 每个兴趣在当月的权重按照时间进行衰减,得到衰减后的兴趣权重,然后对这些衰减后的兴趣权重进行统计,得到用户的长期兴趣,等等。

[0103] 204、推荐服务器根据该用户行为数据确定用户当前点击的兴趣权值,得到用户的实时兴趣。

[0104] 例如,若用户当前点击的消息包括了“NBA”关键字(或标签),则确定用户当前的兴趣为“NBA”,于是,可以计算该“NBA”关键字的权值,以此类推,便可以得到用户的实时兴趣。

[0105] 其中,步骤202、203和204的执行可以不分先后。

[0106] 205、推荐服务器根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣。

[0107] 例如,可以根据预置的策略将用户的短期兴趣、长期兴趣和实时兴趣进行融合,得到用户兴趣。

[0108] 其中,融合的方式可以有多种,比如,可以分别为短期兴趣、长期兴趣和实时兴趣设置权重,然后根据这些权重将短期兴趣、长期兴趣和实时兴趣融合在一起,又比如,还可以为短期兴趣、长期兴趣和实时兴趣之间的关系设置一个函数,然后通过该函数将短期兴趣、长期兴趣和实时兴趣融合在一起,等等,在此不再赘述。

[0109] 206、推荐服务器获取新闻的倒排索引。例如,具体可以如下:

[0110] 从原始实时信息库中获取原始新闻,对获取到的原始新闻进行特征提取,根据提取到的特征对该原始新闻进行分类预测和话题预测,以确定原始新闻的类目和主题;对获取到的原始新闻的内容进行词性加权处理后,进行文本域加权,以确定原始新闻所属的关键词;根据原始新闻的类目、主题和关键词计算该原始新闻库中原始新闻的倒排索引,得到新闻的倒排索引。

[0111] 其中,该原始实时信息库可以存储在该推荐服务器中,也可以存储在其他的设备,比如信息服务器中。

[0112] 其中,可以采用liblinear(一种用来进行分类器生成的技术)来对原始新闻的分类进行预测(即分类预测),得到原始新闻的类目

[0113] 其中,可以采用主题模型(LDA,Latent Dirichlet Allocation)来对原始新闻的主题进行预测(即话题预测),即可以采用LDA来识别原始新闻的文档中潜藏的主题信息,得到原始新闻的主题。

[0114] 其中,可以采用词频-逆向文件频率(TF-IDF,term frequency-inverse document frequency)来对获取到的原始新闻的内容进行词性加权处理,其中,TF-IDF是一种用于信息检索与数据挖掘的常用加权技术,它也是一种统计方法,用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。

[0115] 其中,步骤206与步骤201~205的执行可以不分先后。

[0116] 207、推荐服务器根据该用户兴趣,从新闻的倒排索引中召回相应的新闻,得到候选推荐信息。

[0117] 例如,具体可以根据用户兴趣确定用户所感兴趣的类目、主题和/或关键字,从该新闻的倒排索引中获取与用户所感兴趣的类目、主题和/或关键字相同、相似或相近的原始新闻,得到候选推荐信息。

[0118] 208、推荐服务器基于该候选推荐信息推荐新闻给该用户。例如,具体可以如下:

[0119] A、计算该候选推荐信息中各新闻与该用户兴趣的匹配度,得到新闻与用户的兴趣相关性。

[0120] B、确定该候选推荐信息中各新闻的发布时间确定各新闻的新鲜度,得到新闻的时新性。其中,越新的新闻,其时新性越高。

[0121] C、确定该候选推荐信息中各新闻的点击率,并根据该点击率计算点击模型因子(CM)。其中,点击率越大的新闻,其模型因子越大。

[0122] D、确定该候选推荐信息中各新闻的信息质量。

[0123] 比如,可以通过文本识别的方式确定一个新闻的质量因子,其中,垃圾文章、广告

文章的质量分低。

[0124] E、根据上该兴趣相关性、时新性、点击模型因子和信息质量从该候选推荐信息中确定推荐新闻,比如,具可以如下:

[0125] 根据该兴趣相关性、时新性、点击模型因子和信息质量对该候选推荐信息中的新闻进行评分;将评分高于预置阈值的新闻确定为推荐新闻。

[0126] F、将该推荐新闻推荐给该用户。

[0127] 其中,步骤A、B、C和D的执行步骤可以不分先后。

[0128] 由上可知,本实施例采用获取用户行为数据,根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,然后根据用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,并基于该用户兴趣向用户推荐新闻;由于该方案在计算用户的兴趣时,不仅将用户的长期兴趣作为考量因素,而且,也将用户的短期兴趣与实时兴趣也作为考量因素,以反映出用户的兴趣随时间变化的情况,因此,相对于现在技术而言,可以更加灵活、准确且及时地将用户当前最感兴趣的新闻推荐给用户,在提高时效性的同时,大大改善推荐效果。

[0129] 此外,在计算新闻的过程中,不仅可以考虑新闻与用户的兴趣相关性,同时也可以考虑新闻的时新性,以及关键字的点击率和新闻质量等,因此,可以更加准确的描述这些信息与用户兴趣之间关系强弱,进一步提高了推荐质量和推荐效果。

[0130] 实施例三、

[0131] 为了更好地实施以上方法,本发明实施例还提供一种实时信息的推荐装置,如图3a所示,该实时信息的推荐装置包括获取单元301、运算单元302、确定单元303和推荐单元304,如下:

[0132] (1) 获取单元301;

[0133] 获取单元301,用于获取用户行为数据。

[0134] 其中,用户行为数据指的是可供用户行为分析的相关数据,比如用户的浏览记录、点击记录和/或下载记录等数据。以存储在用户服务器为例,则此时,具体可以从用户服务器中获取用户行为数据。

[0135] (2) 运算单元302;

[0136] 运算单元302,用于根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣。

[0137] 例如,该运算单元302可以包括第一计算子单元、第二计算子单元和第三计算子单元,如下:

[0138] 第一计算子单元,用于根据该用户行为数据计算用户在预置周期中每一天的兴趣权值,得到天兴趣权值,并对该天兴趣权值按照时间进行衰减,得到用户的短期兴趣。

[0139] 例如,该第一计算子单元,具体可以用于根据得到的天兴趣权值确定当前需要进行衰减的兴趣权值,对该需要进行衰减的兴趣权值按照时间进行衰减,得到衰减后的兴趣权值;返回执行根据天兴趣权值确定当前需要进行衰减的兴趣权值的操作,直至该天兴趣权值中所有需要进行衰减的兴趣权重衰减完毕;对得到的所有衰减后的兴趣权值进行统计,得到用户的短期兴趣。

[0140] 其中,在对需要进行衰减的兴趣权值进行衰减时,该第一计算子单元,具体可以用于确定该需要进行衰减的兴趣权值所在的日期与当前日期的日期差;计算该日期差与预置

衰减系数的乘积,并计算1与该乘积的差;将该需要进行衰减的兴趣权值乘以该差,得到衰减后的兴趣权值,具体可参见前面的方法实施例,在此不再赘述。

[0141] 其中,预置周期可以根据实际应用的需求进行设置,比如,一般可以设置为7天、15天或30天等。

[0142] 第二计算子单元,用于根据该用户行为数据计算用户在预置时间范围内的兴趣权值,得到用户的长期兴趣。

[0143] 其中,该预置时间范围可以根据实际应用的需求进行设置,至少大于一天,比如,可以设置为一个季度,一年或两年,等等。

[0144] 例如,以一年为例,则该第二计算子单元,具体可以用于根据该用户行为数据统计用户在当前日期的一年内每个月的用户行为;根据该每个月的用户行为计算每个兴趣在当月的权重;根据该每个兴趣在当月的权重计算一年内每个兴趣的平均权重,对该平均权重进行统计,得到用户的长期兴趣。

[0145] 可选的,第二计算子单元也可以对每个兴趣在当月的权重按照时间进行衰减,得到衰减后的兴趣权重,然后对这些衰减后的兴趣权重进行统计,得到用户的长期兴趣,等等。

[0146] 第三计算子单元,用于根据该用户行为数据确定用户当前点击的兴趣权值,得到用户的实时兴趣。

[0147] (3) 确定单元303;

[0148] 确定单元303,用于根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣。

[0149] 例如,确定单元303,具体可以用于根据预置的策略将用户的短期兴趣、长期兴趣和实时兴趣进行融合,得到用户兴趣。

[0150] 其中,融合的方式可以有多种,比如,可以分别为短期兴趣、长期兴趣和实时兴趣设置权重,然后根据这些权重将短期兴趣、长期兴趣和实时兴趣融合在一起,又比如,还可以为短期兴趣、长期兴趣和实时兴趣之间的关系设置一个函数,然后通过该函数将短期兴趣、长期兴趣和实时兴趣融合在一起,等等,在此不再赘述。

[0151] (4) 推荐单元304;

[0152] 推荐单元304,用于基于该用户兴趣向用户推荐实时信息。

[0153] 其中,该实时信息具体可以为新闻等信息。

[0154] 例如,该推荐单元304可以包括召回子单元和推荐子单元,如下:

[0155] 该召回子单元,用于根据该用户兴趣,从实时信息的倒排索引中召回相应的实时信息,得到候选推荐信息。

[0156] 该推荐子单元,用于基于该候选推荐信息推荐实时信息给该用户。具体可以如下:

[0157] 该推荐子单元,具体可以用于计算该候选推荐信息中各实时信息与该用户兴趣的匹配度,得到实时信息的兴趣相关性;确定该候选推荐信息中各实时信息的发布时间确定各实时信息的新鲜度,得到实时信息的时新性;确定该候选推荐信息中各实时信息的点击率,并根据该点击率计算点击模型因子;根据该兴趣相关性、时新性和点击模型因子从该候选推荐信息中确定推荐信息;将该推荐信息推荐给该用户。

[0158] 可选的,为了提高推荐的信息质量,在确定推荐消息之前,还可以对候选推荐信息中各实时信息的信息质量进行评判,比如,可以通过文本识别的方式确定一个新闻的质量

因子,其中,垃圾文章、广告文章的质量分低。即:

[0159] 该推荐子单元,具体可以用于计算该候选推荐信息中各实时信息与该用户兴趣的匹配度,得到实时信息的兴趣相关性;确定该候选推荐信息中各实时信息的发布时间确定各实时信息的新鲜度,得到实时信息的时新性;确定该候选推荐信息中各实时信息的点击率,并根据该点击率计算点击模型因子;确定该候选推荐信息中各实时信息的信息质量;根据该兴趣相关性、时新性、点击模型因子和信息质量从该候选推荐信息中确定推荐信息;将该推荐信息推荐给该用户。

[0160] 需说明的是,其中,实时信息的倒排索引可以通过对原始实时信息进行收集和统计来得到,即如图3b所示,该实时信息的推荐装置还可以包括倒排索引确定单元305,如下:

[0161] 倒排索引确定单元305,可以用于从原始实时信息库中获取原始实时信息;对获取到的原始实时信息进行特征提取;根据提取到的特征对该原始实时信息进行分类预测和话题预测,以确定原始实时信息类目和主题;对获取到的原始实时信息的内容进行词性加权处理后,进行文本域加权,以确定原始实时信息所属的关键字;根据原始实时信息类目、主题和关键字计算该原始实时信息库中原始实时信息的倒排索引,得到实时信息的倒排索引;

[0162] 则此时,召回子单元,具体用于根据用户兴趣确定用户所感兴趣的类目、主题和/或关键字,从该实时信息的倒排索引中获取与用户所感兴趣的类目、主题和/或关键字相同、相似或相近的原始实时信息,得到候选推荐信息。

[0163] 具体实施时,以上各个单元可以作为独立的实体来实现,也可以进行任意组合,作为同一或若干个实体来实现,以上各个单元的具体实施可参见前面的方法实施例,在此不再赘述。

[0164] 该实时信息的推荐装置具体可以集成在服务器,比如推荐服务器等设备中。

[0165] 由上可知,本实施例的实时信息的推荐装置的获取单元301可以获取用户行为数据,然后由运算单元302根据该用户行为数据分别计算用户的短期兴趣、长期兴趣和实时兴趣,再然后,由确定单元303根据用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣,并由推荐单元304基于该用户兴趣向用户推荐实时信息;由于该方案在计算用户的兴趣时,不仅将用户的长期兴趣作为考量因素,而且,也将用户的短期兴趣与实时兴趣也作为考量因素,以反映出用户的兴趣随时间变化的情况,因此,相对于现在技术而言,可以更加灵活、准确且及时地将用户当前最感兴趣的实时信息推荐给用户,在提高时效性的同时,大大改善推荐效果。

[0166] 此外,在推荐单元304在计算候选推荐信息的过程中,不仅可以考虑实时信息的兴趣相关性,同时也可以考虑实时信息的时新性,以及关键字的点击率和信息质量等,因此,可以更加准确的描述这些信息与用户兴趣之间关系强弱,进一步提高了推荐质量和推荐效果。

[0167] 实施例四、

[0168] 此外,本发明实施例还提供一种实时信息的推荐系统,可以包括本发明实施例提供的任一种实时信息的推荐装置,具体可参见实施例三。其中,该实时信息的推荐装置具体可以集成在服务器,如推荐服务器中,以集成在推荐服务器中,则具体可以如下:

[0169] 推荐服务器,用于获取用户行为数据;根据该用户行为数据分别计算用户的短期

兴趣、长期兴趣和实时兴趣；根据该用户的短期兴趣、长期兴趣和实时兴趣确定用户兴趣；基于该用户兴趣向用户推荐实时信息。

[0170] 其中，计算用户的短期兴趣、长期兴趣和实时兴趣的操作具体可以如下：

[0171] (1) 短期兴趣；

[0172] 用户行为数据计算用户在预置周期中每一天的兴趣权值，得到天兴趣权值，并对该天兴趣权值按照时间进行衰减，得到用户的短期兴趣。

[0173] 其中，预置周期可以根据实际应用的需求进行设置，比如，一般可以设置为7天、15天或30天等。

[0174] (2) 长期兴趣；

[0175] 根据该用户行为数据计算用户在预置时间范围内的兴趣权值，得到用户的长期兴趣。

[0176] 其中，该预置时间范围可以根据实际应用的需求进行设置，至少大于一天，比如，可以设置为一个季度，一年或两年，等等。

[0177] (3) 实时兴趣；

[0178] 根据该用户行为数据确定用户当前点击的兴趣权值，得到用户的实时兴趣。

[0179] 其中，推荐服务器在向用户推荐实时信息时，具体可以根据该用户兴趣，从实时信息的倒排索引中召回相应的实时信息，得到候选推荐信息，然后基于该候选推荐信息推荐实时信息给该用户，比如，可以计算实时信息的兴趣相关性、时新性、点击模型因子和信息质量等参数，然后根据上该兴趣相关性、时新性、点击模型因子和信息质量从该候选推荐信息中确定推荐信息。

[0180] 需说明的是，其中，实时信息的倒排索引可以通过对原始实时信息进行收集和统计来得到，例如，可以如下：

[0181] 推荐服务器，还可以用于从原始实时信息库中获取原始实时信息；对获取到的原始实时信息进行特征提取；根据提取到的特征对该原始实时信息进行分类预测和话题预测，以确定原始实时信息类目和主题；对获取到的原始实时信息的内容进行词性加权处理后，进行文本域加权，以确定原始实时信息所属的关键字；根据原始实时信息类目、主题和关键字计算该原始实时信息库中原始实时信息的倒排索引，得到实时信息的倒排索引。

[0182] 以上各个操作的具体实施可参见前面的方法实施例，在此不再赘述。

[0183] 此外，该实时信息的推荐系统还可以包括其他的设备，比如用户设备，可选的，还可以包括用户服务器和信息服务器等，如下：

[0184] 用户设备，可以用于接收推荐服务器推荐的实时信息。

[0185] 用户服务器，可以用于保存用户行为数据，并向推荐服务器提供用户行为数据。

[0186] 信息服务器，可以用于保存原始实时信息，并向推荐服务器提供保存的原始实时信息。

[0187] 以上各个设备的具体实施，可参见前面的方法实施例，在此不再赘述。

[0188] 由于该实时信息的推荐系统可以包括本发明实施例提供的任一种实时信息的推荐装置，因此，可以实现本发明实施例提供的任一种实时信息的推荐装置所能实现的有益效果，详见前面的实施例，在此不再赘述。

[0189] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可

以通过程序来指令相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:只读存储器(ROM,Read Only Memory)、随机存取记忆体(RAM,Random Access Memory)、磁盘或光盘等。

[0190] 以上对本发明实施例所提供的一种实时信息的推荐方法、装置和系统进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

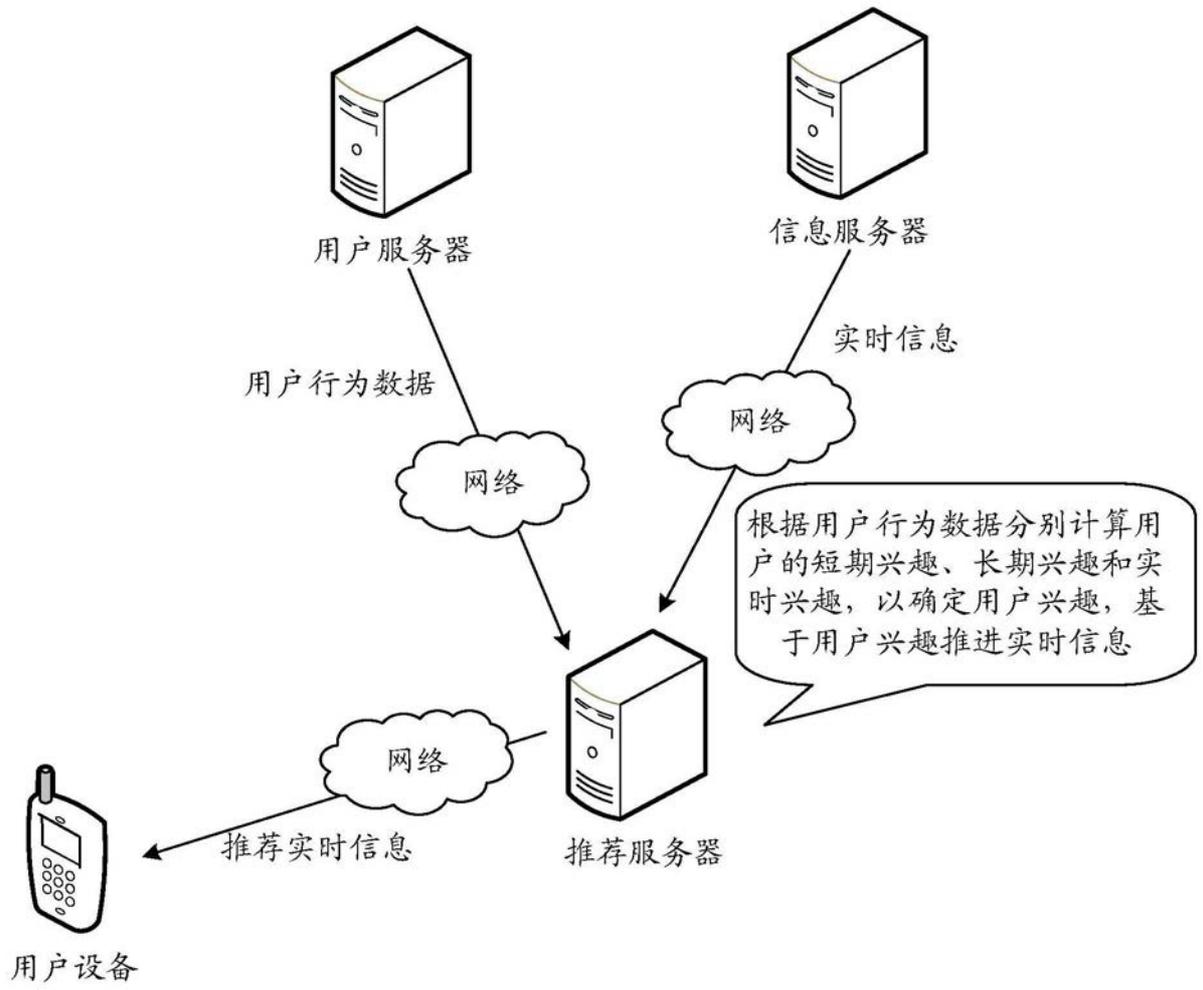


图1a

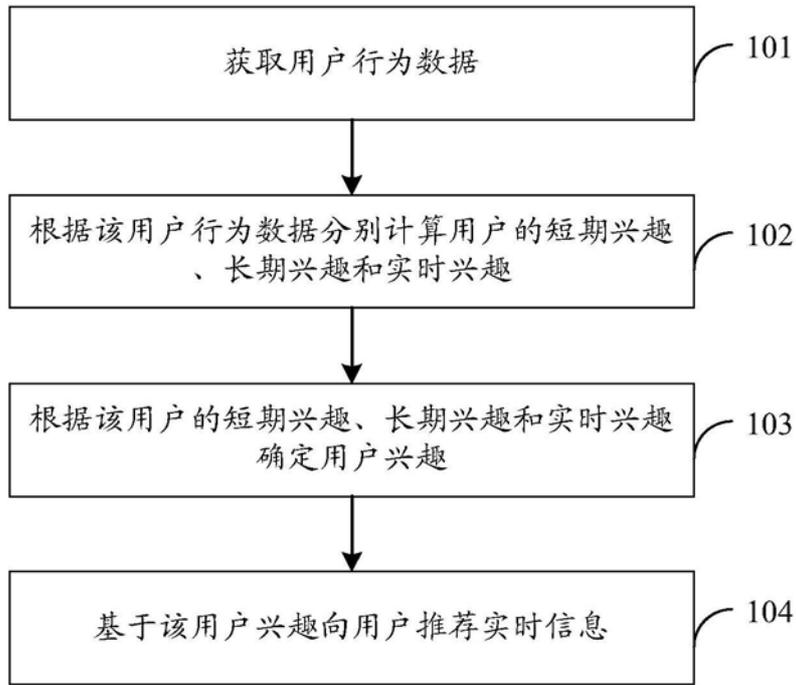


图1b

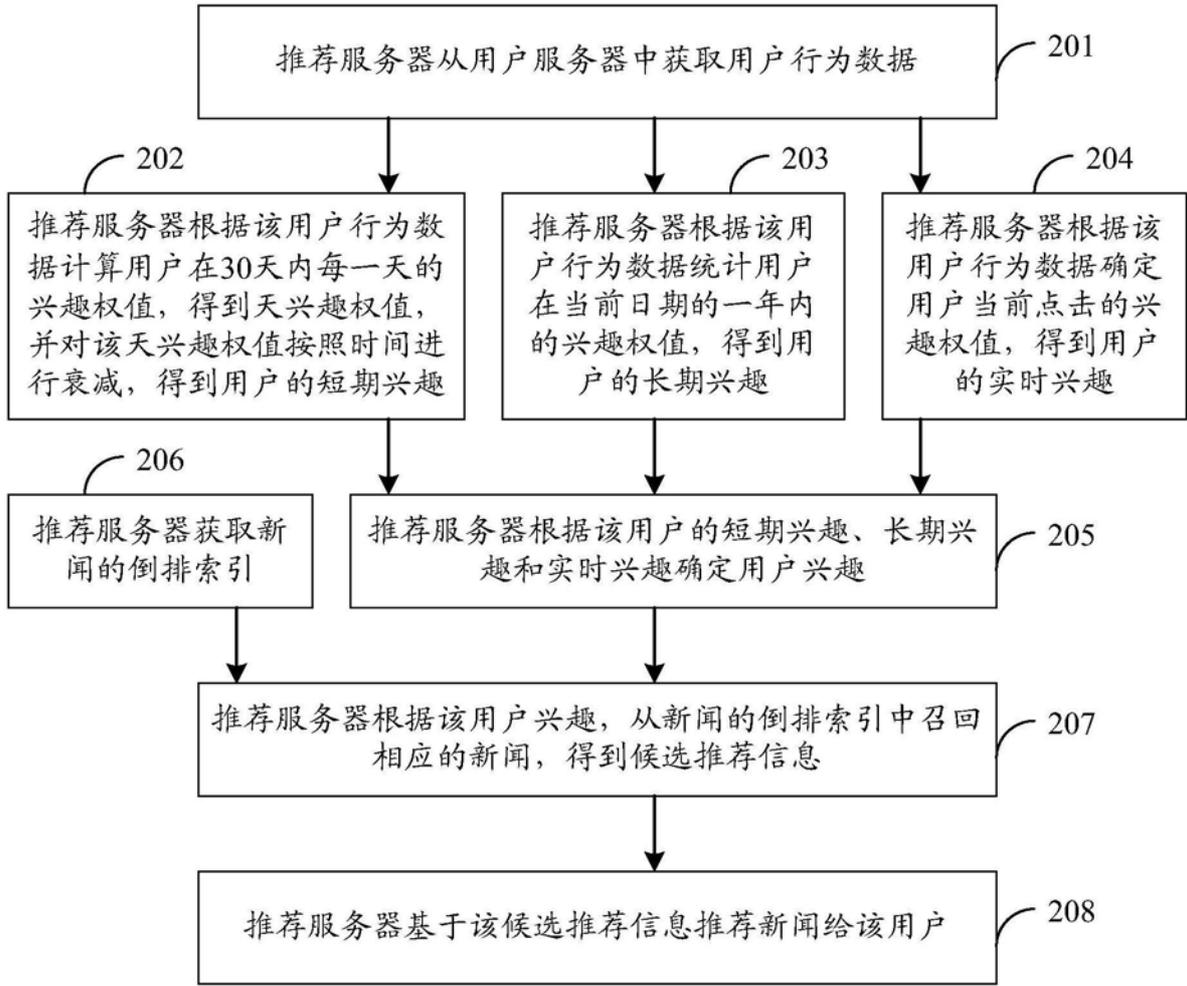


图2

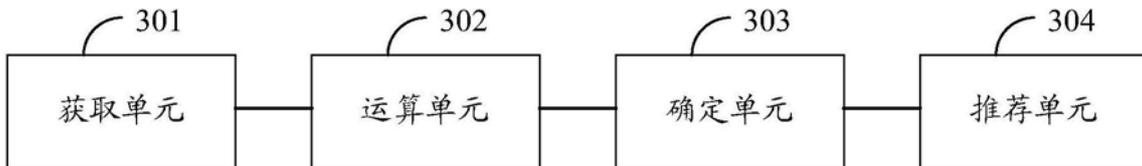


图3a

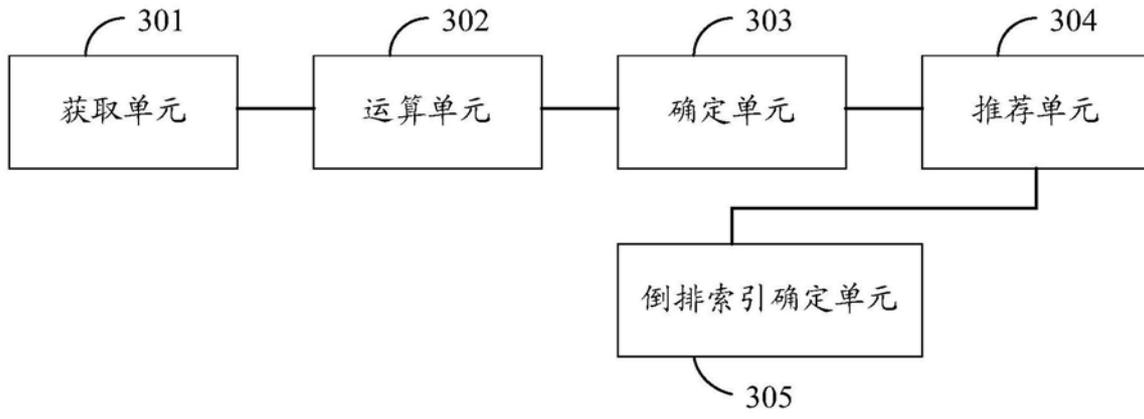


图3b