



(12)发明专利

(10)授权公告号 CN 102956009 B

(45)授权公告日 2017.03.01

(21)申请号 201110235267.5

审查员 孙国辉

(22)申请日 2011.08.16

(65)同一申请的已公布的文献号

申请公布号 CN 102956009 A

(43)申请公布日 2013.03.06

(73)专利权人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼岛资本大厦一座
四层847号邮箱

(72)发明人 顾湘余 傅其乐 祖仲林 宁伟

(74)专利代理机构 北京集佳知识产权代理有限公司

11227

代理人 逯长明 王宝筠

(51)Int.Cl.

G06Q 30/02(2012.01)

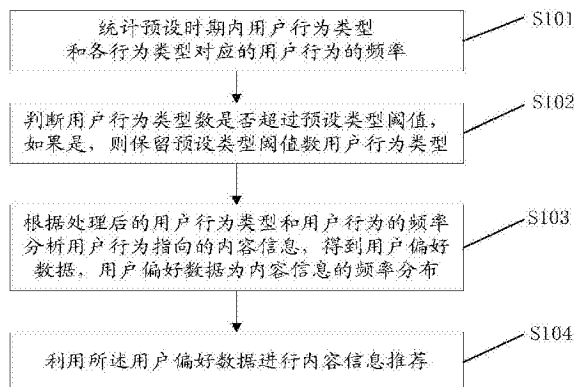
权利要求书3页 说明书12页 附图4页

(54)发明名称

一种基于用户行为的电子商务信息推荐方法与装置

(57)摘要

本申请提供了一种基于用户行为的电子商务信息推荐方法。该方法包括：统计预设时期内用户行为类型和各行为类型对应的用户行为的频率；判断用户行为类型数是否超过预设类型阈值，如果是，则保留预设类型阈值数用户行为类型；根据处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息，得到用户偏好数据，用户偏好数据为内容信息的频率分布；利用所述用户偏好数据进行内容信息推荐。本申请还提供了一种基于用户行为的电子商务信息推荐装置。本申请对用户行为数据进行优化，剔除了可能存在的“伪数据”，使得在该数据基础上分析得到的用户偏好更加准确真实，处理速度得以提高，推荐效果更好。



1. 一种基于用户行为的电子商务信息推荐方法,其特征在于,该方法包括:

统计预设时期内的用户行为类型和各行为类型对应的用户行为的频率;

采用如下三种实现方式中的任意一种得到用户偏好数据,所述用户偏好数据为各内容信息的频率分布:

第一种实现方式:判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型,根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据;

第二种实现方式,判断各用户行为的频率是否超过预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率,根据用户行为类型和上述处理后用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据;

第三种实现方式,判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型,判断各用户行为的频率是否超过预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率,根据上述处理后的用户行为类型和上述处理后用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据;

利用所述用户偏好数据向用户推荐内容信息。

2. 根据权利要求1所述的方法,其特征在于,在分析用户行为指向的内容信息前,所述方法还包括:对用户行为进行时间影响消除处理,该处理步骤包括:对用户行为赋予时间权重值,所述时间权重值为越靠近预设时期结束时刻权重值越高,越靠近预设时期的起始时刻权重值越低。

3. 根据权利要求2所述的方法,其特征在于,所述时间权重值按照下述公式得到:

$$\text{时间权重值} = 1/\ln(D_{\text{off}})$$

式中 D_{off} 为用户行为发生时刻到预设时期结束时刻的时间差。

4. 根据权利要求1所述的方法,其特征在于,在分析用户行为指向的内容信息前,所述方法还包括:对内容信息进行标准化处理,该处理步骤包括:根据标准产品单元和行业知识库校正内容信息。

5. 根据权利要求1所述的方法,其特征在于,所述分析用户行为指向的内容信息得到用户偏好数据的步骤包括:

根据用户行为的发生时间将处理后的用户行为类型和用户行为的频率在预设时期内分成至少两个时段单元;

根据各时段单元内的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到该时段单元的用户偏好数据;

根据各时段单元的用户偏好数据过滤掉迁移性、突发性偏好数据以得到整个预设时期的用户偏好数据。

6. 根据权利要求1所述的方法,其特征在于,所述内容信息包括产品信息时,则所述分析用户行为指向的内容信息得到用户偏好数据的步骤包括:

判断产品标定价格是否在预设价格范围内,如果否,则将产品价格设置在预设价格范围内;和/或,判断产品的所属行业与提供该产品的供应商的注册行业是否一致,如果否,则将该产品信息滤除。

7. 根据权利要求1所述的方法,其特征在于,利用用户偏好数据向用户推荐内容信息包

括：

选取预设数量的其他用户组成参考用户集，计算待推荐用户与参考用户集内各用户的相似度，将相似度超过预设相似度阈值的用户组成相似用户集；

分析相似用户集内各用户的用户行为得到该相似用户集的群体偏好数据，所述群体偏好数据为用户行为指向的内容信息的频率分布；

根据待推荐用户的用户偏好数据和所述群体偏好数据向待推荐用户推荐内容信息。

8. 根据权利要求7所述的方法，其特征在于，所述选取预设数量的其他用户组成参考用户集包括：

判断待推荐用户的IP地址对应的地域与该用户注册的地域是否一致，如果是，则选取待推荐用户注册地域地的其他用户组成参考用户集；如果否，则选取待推荐用户的IP地址对应的地域地的其他用户组成参考用户集。

9. 根据权利要求1所述的方法，其特征在于，利用用户偏好数据向用户推荐内容信息包括：

选取预设数量的用户组成规则创建用户集，分析所述规则创建用户集内各用户的用户行为得到各用户的用户偏好数据，根据各用户的用户偏好数据训练得到规则集，所述规则集为用以预测内容信息推荐的各类用户偏好的组合规则集；

将待推荐用户的用户偏好数据与所述规则集进行匹配，根据匹配成功的偏好数据向待推荐用户推荐内容信息。

10. 根据权利要求1所述的方法，其特征在于，当用户行为类型数超过预设类型阈值时，保留的用户行为类型包括：频率小于等于预设频率阈值的用户行为对应的用户行为类型。

11. 一种基于用户行为的电子商务信息推荐装置，其特征在于，该装置包括：

统计模块、判断模块、偏好分析模块和信息推荐模块，其中：

所述统计模块，用于统计预设时期内的用户行为类型和各行为类型对应的用户行为的频率；

所述判断模块和所述偏好分析模块采用如下三种结构中的任意一种结构：

第一种结构：所述判断模块，用于判断用户行为类型数是否超过预设类型阈值，如果是，则保留小于等于预设类型阈值数的用户行为类型，

所述偏好分析模块，用于根据上述处理后的用户行为类型分析用户行为指向的内容信息，得到用户偏好数据；

第二种结构：所述判断模块，用于判断各用户行为的频率是否超过预设频率阈值，如果是，则将预设频率阈值作为该用户行为的频率，

所述偏好分析模块，用于根据上述处理后用户行为的频率分析用户行为指向的内容信息，得到用户偏好数据；

第三种结构：所述判断模块，用于判断用户行为类型数是否超过预设类型阈值，如果是，则保留小于等于预设类型阈值数的用户行为类型，判断各用户行为的频率是否超过预设频率阈值，如果是，则将预设频率阈值作为该用户行为的频率，所述偏好分析模块，根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息，得到用户偏好数据；

所述信息推荐模块，用于利用用户偏好数据向用户推荐内容信息。

12. 根据权利要求11所述的装置,其特征在于,所述装置还包括时间影响消除模块,用于在分析用户行为指向的内容信息前,对用户行为赋予时间权重值,所述时间权重值为越靠近预设时期结束时刻权重值越高,越靠近预设时期起始时刻权重值越低。

13. 根据权利要求11所述的装置,其特征在于,所述偏好分析模块包括:

时段单元划分子模块、偏好分析子模块和过滤子模块,其中:

所述时段单元划分子模块,用于根据用户行为的发生时间将处理后的用户行为类型和用户行为的频率在预设时期内划分成至少两个时段单元;

所述偏好分析子模块,用于根据各时段单元内的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到该时段单元的用户偏好数据;

所述过滤子模块,用于根据各时段单元的用户偏好数据过滤掉迁移性、突发性偏好数据以得到整个预设时期的用户偏好数据。

14. 根据权利要求11所述的装置,其特征在于,所述装置还包括相似度计算模块,用于在选取预设数量的其他用户组成参考用户集后,计算待推荐用户与参考用户集内各用户的相似度,将相似度超过预设相似度阈值的用户组成相似用户集;

所述偏好分析模块还用于分析相似用户集内各用户的用户行为得到该相似用户集的群体偏好数据,所述群体偏好数据为用户行为指向的内容信息的频率分布;

所述信息推荐模块,还用于根据待推荐用户的用户偏好数据和所述群体偏好数据向待推荐用户推荐内容信息。

15. 根据权利要求11所述的装置,其特征在于,所述装置还包括规则集创建模块,用于在选取预设数量的用户组成规则创建用户集,调用偏好分析模块分析所述规则创建用户集内各用户的用户行为得到各用户的用户偏好数据,根据各用户的用户偏好数据训练得到规则集,所述规则集为用以预测内容信息推荐的各类用户偏好的组合规则集,则:

所述偏好分析模块还用于分析规则创建用户集内各用户的用户行为得到各用户的用户偏好数据;

所述信息推荐模块,还用于将待推荐用户的用户偏好数据与所述规则集进行匹配,根据匹配成功的用户偏好数据向待推荐用户推荐内容信息。

一种基于用户行为的电子商务信息推荐方法与装置

技术领域

[0001] 本申请涉及电子商务信息推荐技术领域,尤其涉及一种基于用户行为的电子商务信息推荐方法与装置。

背景技术

[0002] 随着互联网技术的发展,电子商务蓬勃兴起。电子商务是利用计算机技术、网络技术和远程通信技术,使商务交易中的买卖双方在不需谋面的情况下实现各种商贸活动。要实现商贸交易,买方必须预先获知商品的内容信息以便决策是否购买该商品。商品内容信息包括商品所属行业类目、商品的提供商、商品价格以及与该商品相关的资讯信息等。在电子商务信息交易平台上,获知所述内容信息的方式主要有两种:一种为用户(买方)主动搜索方式,即用户在明确自己的购买需求后,将需求以关键词的形式体现出来,然后利用该关键词在海量的多样性数据中检索,从而获得需要的内容信息;另一种是用户被动接受型,即卖方通过电子商务交易平台向用户推荐商品的内容信息,用户被动接受卖方推荐的商品内容信息后,在该内容信息的导引下购买相关产品。对于第二种方式,商贸交易的卖方为了提高交易成功率,往往不会将自己的所有信息直接推荐给用户,而是首先分析用户过去的行为建立用户偏好数据,然后根据用户偏好针对性地推荐用户可能感兴趣的特定信息。这种推荐方式在电子商务信息交易中能显著提高用户体验,增加卖家曝光的精准度,能有效引导用户迅速成为买家,从而降低交易成本。但是,电子商务信息的信息推荐者进行用户偏好分析时获取的用户历史行为数据可能存在偏差、错误,甚至包括以用户名义恶意引导消费倾向的作弊数据,建立在这些数据基础上的分析结果必然不能真实地反映用户偏好特点,从而导致推荐内容信息偏离用户需求,推荐效果受到影响。另外,由于电子商务网站的信息访问量特别巨大,电子商务网站服务器中存在海量的用户行为数据,针对海量用户行为数据分析得到用户偏好数据对推荐系统自身处理能力是一种严峻考验,海量的用户行为数据减缓了推荐系统的分析处理速度。

发明内容

[0003] 有鉴于此,本申请的发明目的在于提供一种基于用户行为的电子商务信息推荐方法与装置,通过对用户行为数据的优化,一方面滤除可能存在的“伪数据”以校正用户偏好数据,从而确保根据用户偏好数据推荐的内容信息符合用户的真实偏好;另一方面减少用于用户偏好分析的数据量以提高偏好数据的分析速度,从而确保用户快速的获取到推荐的内容信息。

[0004] 本申请提供的基于用户行为的电子商务信息推荐方法包括:

[0005] 统计预设时期内用户行为类型和各行为类型对应的用户行为的频率;

[0006] 判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型;和/或,判断各用户行为的频率是否超过预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率;

[0007] 根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据,所述用户偏好数据为各内容信息的频率分布;

[0008] 利用用户偏好数据向用户推荐内容信息。

[0009] 本申请还提供了一种基于用户行为的电子商务信息推荐装置,该装置包括:

[0010] 统计模块、判断模块、偏好分析模块和信息推荐模块,其中:

[0011] 所述统计模块,用于统计预设时期内的用户行为类型和各行为类型对应的用户行为的频率;

[0012] 所述判断模块,用于判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型;和/或,判断各用户行为的频率是否超过预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率;

[0013] 所述偏好分析模块,用于根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据,所述用户偏好数据为各内容信息的频率分布;

[0014] 所述信息推荐模块,用于利用用户偏好数据向用户推荐内容信息。

[0015] 本申请的技术方案通过判断用户行为类型数与用户行为的频率是否符合预设阈值的方式对用户的历史行为数据进行优化,在该优化处理后的用户行为数据基础上分析得到用户偏好,然后根据该偏好向用户推荐相应的内容信息。与现有技术相比,本申请的优化处理过程排除了用户行为数据中的恶意行为、作弊行为等“伪数据”,使得根据该优化数据进行的偏好分析结果更真实地反映了用户的偏好特点,进而使得根据该用户偏好进行的信息推荐的推荐效果更好,提高了电子商务信息的交易成功率。另外,由于根据偏好分析结果推荐的信息能够符合用户的实际需求愿望,这样用户只需根据电子商务交易平台推荐的信息完成交易即可,而不必再反复查找其所需的内容信息,降低了反复查找内容信息给电子商务平台带来的访问压力。并且,由于本申请的优化处理过程去除了用户行为数据中的大量“伪数据”,减少了用于信息推荐分析的数据量,提高了信息推荐分析的速度,能够更加快速的获取到推荐信息。

附图说明

[0016] 图1为本申请的方法的实施例一的流程图;

[0017] 图2为本申请的方法的实施例二的流程图;

[0018] 图3为本申请的方法的实施例三的流程图;

[0019] 图4为本申请的方法的实施例四的流程图;

[0020] 图5为本申请的方法的实施例五的流程图;

[0021] 图6为本申请的方法的实施例六的内容信息推荐的流程图;

[0022] 图7为本申请的方法的实施例七的内容信息推荐的流程图;

[0023] 图8为本申请的装置的实施例八的组成框图。

具体实施方式

[0024] 本申请通过统计用户行为类型和用户行为类型对应的用户行为的频率,将该统计结果与预设阈值进行比较,以滤除用户行为数据中的恶意行为、作弊行为等“伪数据”,然后

在上述经过优化处理后的用户行为数据基础上进行用户偏好分析,根据分析得到的用户偏好特点实现相应内容信息的推荐。

[0025] 为使本领域技术人员进一步了解本申请的特征及技术内容,下面结合附图和实施例,对本申请的技术方案进行详细描述。

[0026] 实施例一

[0027] 电子商务信息交易平台是虚拟的市场,买卖主体不需要面对面磋商即可完成交易。交易的成功与否取决于交易中买方的求购愿望与卖方的供给内容间的匹配程度。卖方主动提供商品信息固然可以增加卖家的曝光率,提高交易成功概率,但是,如果卖方主动供给内容与买方需求偏差太大,卖方即便使出“浑身解数”,可能仍将甚微收获。因此,为提高交易成功率,卖家的主动推送行为往往需要建立在对用户的分析基础上,这种分析需要预先收集大量的用户信息,包括用户的历史行为数据、用户属性数据等内容,利用这些用户训练得到某个内容信息的倾向预测数据,进而将某个特定用户的用户偏好与该倾向预测数据进行匹配以决定如何向用户推荐相应的内容信息。分析用户偏好通常建立在用户的历史行为数据基础上,用户历史行为反映了用户的求购愿望。比如,某个用户行为在一段时间内集中高频出现,该行为指向的内容信息也集中出现在某些商品,则说明该用户在该时期内的偏好就是这些商品,该用户有极大可能会购买该商品,如果信息提供商能够提供与该偏好匹配的产品信息,必然增大了交易成功率。

[0028] 然而,正如背景技术所言,有些用户的历史行为数据可能不真实,是“伪数据”,比如现实中可能存在这样的情形:在某个时期内大部分用户均关注某一产品,则说明该产品是该时段内的新潮产品,由于用户的个人偏好大多受大众消费趋势的影响,那么单个用户的偏好将转向该商品信息,从而导致从事该商品交易的卖家利润可观。基于利润的争夺,有些卖家便抓住用户的这种心理,以用户名义在某时期内大量重复某些行为,以图诱导消费趋势。这样用户的历史行为数据中将出现“作弊数据”,建立在这些数据基础上的用户偏好分析将“失真”,进而导致推荐的内容信息与用户的真实需求偏离,降低交易成功率。另外,由于分析用户偏好需要建立在海量的用户的历史行为数据基础上,再加上某些卖家恶意以用户名义在某时期内大量重复某些行为,更进一步的增加了进行用户偏好分析所依据的历史行为数据量,导致用户偏好的分析速度缓慢。本申请的实施例提供了一种基于用户行为的电子商务信息推荐方法,该方法对用户历史行为数据进行优化处理以滤除“伪数据”。参见附图1,本实施例提供的一种基于用户行为的电子商务信息推荐方法包括:

[0029] 步骤S101:统计预设时期内用户行为类型和各行为类型对应的用户行为的频率;

[0030] 获取用户的偏好数据需要根据用户的历史行为进行分析,分析的前提是要选择一个参考时间段,即设定一个预设时期,该预设时期的长短将影响到用户偏好的分析结果:预设时期长度不同,选择的用户行为类型的数量和用户行为的频率便不同,由此得出的用户偏好也可能大不相同,一般而言,该预设时期不宜选择过长,过长则无法准确反应用户偏好的迁移特性和细微变化,也不宜选择过短,过短则可能导致用户的偏好不具有代表性;预设时期确定后,将统计分析该预设时期内的用户行为类型以及该行为类型对应的用户行为的频率,这里的用户行为类型包括但不限于检索、浏览、点击、保存行为,用户行为的频率是某种用户行为类型在该预设时期内发生的次数,用户行为类型与用户行为频率对应;为后面更见形象的说明起见,此处假设在预设时期T内统计得到五种行为类型,分别为act1、act2、

act3、act4、act5,每种行为类型对应的用户行为的频率分别为 f_1 、 f_2 、 f_3 、 f_4 、 f_5 ,由于“伪数据”的存在,上述五种用户行为类型可能是虚假的,不代表用户真实行为。

[0031] 步骤S102:判断用户行为类型数是否超过预设类型阈值,如果是,则保留预设类型阈值数的用户行为类型;

[0032] 上面曾提到用户行为可能包含有作弊行为,这些作弊行为在建立用户行为偏好数据时需要予以剔除,否则将影响到根据偏好数据进行推荐的推荐效果;进行作弊数据剔除的依据是预设类型阈值,该预设阈值可以通过分析一定时期内大量的用户行为得出的统计数据,也可以是根据实际操作情况得出的一个经验估计值,在此基础上根据推荐效果进行修正调整;上述举例中,假设预设类型阈值 $\theta_{actth}=4$,那么上述五种行为类型就超过了该预设阈值指定的数量,说明该用户行为类型中包含有“伪行为”,这时则将上述的五种用户行为类型按照预设类型阈值数进行保留,即只留下四种用户行为类型,该四种用户行为类型分别为act2、act3、act4、act5,由于用户行为类型与用户行为的频率相联系,因而用户行为的频率也保留下 f_2 、 f_3 、 f_4 、 f_5 四个对应的频率值,需要说明的是这里过滤掉的是act1行为类型,实际上,也可以过滤掉其他行为类型,只要剩余的用户行为类型数量不超过预设类型阈值即可;然而,尽管这种滤除操作并没限定滤除对象,但在一些情况下结合其他因素进行滤除则更为合理,比如,在用户行为类型数超过预设类型阈值时,优先考虑滤除时间最早的行为类型,因为该类型对用户偏好的影响可能已经“过时”,也可以优先考虑将频率最高的用户行为对应的用户行为类型滤除,因为该频率最高的用户行为可能恰巧代表的是作弊行为。该优选实施例的推广应用即为:在用户行为类型数超过预设类型阈值时,保留低于预设频率阈值的用户行为对应的用户行为类型,从而使得用户行为类型数不超过预设类型阈值,通过这两种方式可以有效滤除作弊行为。

[0033] 步骤S103:根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据,所述用户偏好数据为各内容信息的频率分布;

[0034] 用户的每个行为均对应一个或者多个行为对象,该行为对象即是用户行为指向的内容信息,该内容信息可以是单个信息,也可以是复合信息,比如用户在网络中搜索了一件商品,用户行为类型则为“搜索行为”,该“搜索行为”指向的内容信息即是该商品,该商品的属性可以包括商品价格、商品形状、商品所属行业类目等中的一种,也可以是上述商品属性中的几种,甚至还可进一步延伸包括商品提供商、提供商的资质情况等;在上述的举例中,假设上述每种用户行为类型指向的内容信息分别是:act2指向三种内容信息C1、C2、C3,这三种内容信息出现的频率分别为 f_{21} 、 f_{22} 、 f_{23} ,此三种频率的总和为 f_2 ,表示为act2 (C1 [f_{21}]、C2 [f_{22}]、C3 [f_{23}]);act3 (C1 [f_{31}]、C2 [f_{32}]、C3 [f_{33}]、C4 [f_{34}]、C5 [f_{35}])、act4 (C2 [f_{42}]、C3 [f_{43}]、C5 [f_{45}])、act5 (C2 [f_{52}]、C3 [f_{53}]、C4 [f_{54}]、C5 [f_{55}]);得到上述数据后,以用户行为指向的内容信息为依据绘制用户偏好数据曲线,将该曲线顶点(频率值最大)对应的内容信息作为用户偏好:上述数据共出现C1、C2、C3、C4、C5五种内容信息,他们的频率值分别为 $f_{21}+f_{31}$ 、 $f_{22}+f_{32}+f_{42}+f_{52}$ 、 $f_{23}+f_{33}+f_{43}+f_{53}$ 、 $f_{34}+f_{54}$ 、 $f_{35}+f_{45}+f_{55}$,选取其中频率值最高的对应的内容信息作为该用户的用户偏好数据,也可以先设定一个预设值,将超过该预设值的对应的多个内容信息作为该用户的用户偏好数据。

[0035] 步骤S104:利用用户偏好数据向用户推荐内容信息。

[0036] 上述步骤中已经分析出了用户的偏好数据,则根据该偏好数据即可进行内容推

荐。

[0037] 本申请的实施例通过将用户的历史行为类型和该类型对应的频率与预设阈值进行比较,排除用户历史行为中的作弊数据、恶意数据,在该处理后的数据基础之上进行用户偏好分析,根据分析得到的用户偏好数据实现内容信息的推荐。与现有技术相比,本实例建立用户偏好数据时对用户行为数据进行了优化处理,消除了可能的“伪数据”,使得根据较为准确的用户行为数据得到的用户偏好能真实地反映用户的实际需求愿望,进而获得较好的推荐效果,提高电子商务信息交易的成功率。实际应用中,由于推荐的商品信息能够符合用户的实际需求愿望,这样用户只需根据电子商务交易平台推荐的信息完成交易即可,而不必再反复查找其所需的商品信息,降低了反复查找商品信息给电子商务平台带来的访问压力。另外,由于本申请的优化处理过程去除了用户行为数据中的大量“伪数据”,减少了用于信息推荐分析的数据量,提高了信息推荐分析的速度,能够更加快速的获取到推荐信息。

[0038] 上述实施例中,仅仅从用户行为类型的角度对用户行为数据进行了优化,实际上,还可以对用户行为的频率进行优化处理,因为在预设时间段内某个用户行为的频率也可能出现异常,比如“畸高”或“畸低”,这些情形可能反映出该数据是作弊数据,也应当予以剔除。因此,上实施例的步骤S102之后还可以增加下述步骤:

[0039] 判断各用户行为的频率是否超过该用户行为的预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率;

[0040] 经过上述步骤后的用户行为类型和行为频率更加得到优化,建立在此基础上的用户偏好更加准确进而取得更好的推荐效果。值得注意的一点是,上述步骤除可以和实施例一中步骤S102以并列方式进行外,还可以替代步骤S102,同样能达到改善推荐效果的目的。

[0041] 实施例二

[0042] 上述实施例虽然通过与预设类型(频率)阈值的匹配滤除了用户行为中的“伪数据”,经过滤除处理后的用户行为数据已能较好地反映用户的真实偏好。但是,在预设时期T内用户行为的发生时间往往不同,发生时间的差异对分析用户偏好的影响也不同,一般而言,越是靠近该预设时期T的结束时刻发生的用户行为越接近用户的真实偏好,越是靠近该预设时期T的起始时刻发生的用户行为越偏离用户的真实偏好。因此,有必要在上述实施例的基础上对用户行为数据进行时间影响消除处理,实现这种处理的具体方式很多,只要能反映出发生时间靠后的用户行为对用户偏好建立的影响大,发生时间靠前的用户行为对用户偏好建立的影响小,即可实现本申请的发明目的。在实施例一的基础上增加上述处理步骤后的方法构成本申请的又一个实施例,参见附图2,该实施例与实施例一除在步骤S102、S103之间添加下述内容外,其他步骤相同。本实施例给出的时间影响消除处理的步骤包括:

[0043] 对用户行为进行时间影响消除处理,该处理步骤包括:对用户行为赋予时间权重值,所述时间权重值为越靠近预设时期结束时刻权重值越高,越靠近预设时期的起始时刻权重值越低。

[0044] 本实施例对用户行为数据进行了时间影响消除处理,使得以用户行为为基础构建用户偏好的准确度进一步提高,进而推荐效果得到更好改善。

[0045] 尽管时间影响消除的具体实现方式较多,但本申请的实施例优选按照如下的方式计算时间权重值:

[0046] 时间权重值 $=1/\log(D_{\text{off}})$

[0047] 式中 D_{off} 为用户行为发生时刻到预设时期结束时刻的时间差。

[0048] 用户行为被赋予权重值后,用户行为的频率值乘以相应的时间权重值以修正用户行为频率,按照这种方式优化后的用户行为数据将更能准确地反映用户在预设时期内的偏好特征。

[0049] 实施例三

[0050] 用户行为指向的内容信息是分析用户偏好的基础,但是,用户在进行某些用户行为时指向的内容信息受到用户知识能力的限制,可能并不准确,由此可能导致在此基础上的分析出现偏差。比如,用户实施了一个搜索行为,该搜索行为指向“关键词1”,这个“关键词1”是用户根据自己掌握的知识所确定的关键词,但该关键词可能并不准确,甚至有歧义、错误,此刻则需进行“偏差校正”,校正后的数据才能用于进行用户偏好分析。在实施例一的基础上增加上述处理步骤后的方法可构成本申请的再一个实施例,参见附图3,该实施例与实施例一除在步骤S102、S103之间添加下述内容外,其他步骤相同。本实施例给出的内容信息偏差校正方法包括:

[0051] 对内容信息进行标准化处理,该处理步骤包括:根据标准产品单元和行业知识库校正内容信息。

[0052] 标准产品单元SPU (Standard Product Unit) 是一组可复用、易检索的标准化信息的集合,该集合描述了“产品”的特性,存储了产品的基本信息;行业知识库是某行业的“知识”的集合,这些“知识”包括行业标准、行业业务模型、行业业务数据与信息模型、行业信息化全景图、行业领域构件、行业采购链、行业上下游产品等内容,该知识库涵盖的内容相当丰富。用户行为指向的内容信息提取出来后,将其与标准产品单元和行业知识库进行匹配比较,如果两者不一致,则用标准产品单元和行业知识库中的标准语言描述该内容信息,经过这样标准化处理后的内容信息更加便于后续步骤的识别和处理,对于快速、准确地完成用户偏好分析具有良好效果。

[0053] 实施例四

[0054] 前述实施例选取了一个参考时间周期即预设时间 T ,通过对该时期内的用户行为类型和用户行为的频率进行分析得到用户偏好数据。但是,用户在该时间周期内可能发生偏好迁移(变化),也就是说,用户可能在一个时段内对内容信息1感兴趣,在另一个时段内可能对与内容信息1相关联的内容信息2感兴趣(内容信息1与内容信息2形成迁移对),这种情形即是用户偏好的迁移特性,在分析具有偏好迁移特性的用户行为时,需要注意的是:只有迁移对的迁移终点对应的内容信息(内容信息2)才真实地代表用户当下的兴趣,对于迁移对的迁移起点对应的内容信息(内容信息1)表示用户已经“放弃”该内容信息,对此不再感兴趣,因此,应当将迁移终点对应的内容信息作为用户偏好,并据此进行内容信息的推荐。除用户偏好发生迁移的情形外,在预设时期内的某个时段内可能存在用户对某个内容信息极度感兴趣,用户行为频率极大,这种情形构成用户偏好的突发特性,该特性说明该用户的偏好并不稳定,仅代表该用户的激情偏好,而不能代表整个预设时期的偏好,因此也应当在分析用户偏好时予以剔除。与用户偏好的突发特性相对应的是用户偏好的持续特性,即在一个时段内用户的偏好集中在某几个内容信息上,在另一个时段内用户偏好仍然集中在该几个内容信息上,这种情形说明该用户的偏好较为稳定,这些用户行为数据则能较好地反映用户偏好,根据此类用户偏好进行的内容信息推荐有极大的可能符合该用户的需

求,进而提高交易成功率。由此,本申请给出了又一个实施例,该实施例中增加将上述突发性、迁移性用户偏好过滤掉、保留下持续性用户偏好。本实施例除将实施例一的步骤S103变化为下述内容外,其他步骤相同。参见附图4,本实施例给出的分析得到用户偏好数据的步骤包括:

[0055] 步骤A:根据用户行为发生的时间将处理后的用户行为类型和用户行为频率在预设时期内分成至少两个等长时段单元;

[0056] 预设时期选取的时段长度越短,用户偏好的细微程度越大,通过比较不同时段的用户偏好曲线,即可得到突发性用户偏好和迁移性偏好;将预设时期T划分成至少两个等长时段单元,根据实际的需要可选择两个或多个时段单元,时段单元越多,得到的用户偏好变化情,也越细微,这里为方便起见,以两个时段单元为例: T_1 、 T_2 , $T_1+T_2=T$ 。

[0057] 步骤B:根据各时段单元内的用户行为类型和用户行为频率分析用户行为指向的内容信息,得到该时段单元的用户偏好数据;

[0058] 划分时段单元后,分别分析 T_1 、 T_2 时段单元的用户行为类型和用户行为频率分析用户行为指向的内容信息,分析过程可以采取绘制曲线的方式进行:以用户内容信息为横轴、用户内容信息出现的频率为纵轴。

[0059] 步骤C:将各时段单元的用户偏好数据进行比对以滤除迁移性用户偏好数据对中的迁移起点数据和/或突发性用户偏好数据,得到整个预设时期的用户偏好数据;

[0060] 得到 T_1 、 T_2 这两个时段的用户偏好曲线图后,将两个曲线图进行比对分析,将两个时段中均具有的内容信息作为持续性偏好数据予以保留,将仅在前一个时段单元中出现的内容信息作为迁移性用户偏好的迁移起点予以滤除,将仅在后一个时段单元中出现的内容信息作为迁移性用户偏好的迁移终点予以保留,将在任何一个时段单元中出现的突发性用户偏好数据予以滤除,进而得到整个预设时段T的用户偏好数据。

[0061] 本实施例通过将预设时期进行细分,然后比对细分后的每个时段单元的用户偏好数据的变化,滤除了迁移性、突发性用户偏好数据。这样得到的用户偏好更为准确地反映用户在整个预设时期内的用户偏好,进而在此用户偏好数据基础上进行的信息内容推荐的推荐效果更好。

[0062] 实施例五

[0063] 上述实施例为得到较为准确的用户行为数据,进行了一系列的优化操作,滤除了用户行为中的“伪数据”,为分析得到准确的用户偏好奠定了较好的基础。但是,对于用户行为指向的内容信息是否准确还待研究,因为分析用户偏好的落脚点是内容信息,而这些内容信息由电子商务信息卖家提供,卖家为争夺市场利益、抢占用户,也会恶意地在内容信息上制造“伪数据”,比如,当内容信息包含产品价格信息时,卖家可能恶意标定低价以打击对手,或者在垄断情况下随意标定高价以攫取更多利润,这些行为扰乱了电子商务信息交易市场,基于这些“伪”的内容信息进行用户偏好分析得到的结果可能大为“失真”。为此,有必要对内容信息进行“打假”。在实施例一的基础上增加上述处理步骤后的方法可构成本申请的又一个实施例,参见附图5本实施例与实施例一除在步骤S102、S103之间添加下述内容外,其他步骤相同。本实施例给出的排除上述“伪数据”的方法包括:

[0064] 判断产品标定价格是否在预设价格范围内,如果否,则将产品价格设置在预设价格范围内;和/或,判断产品的所属行业与提供该产品的供应商的注册行业是否一致,如果

否,则将该产品信息滤除。

[0065] 经过本实施例对内容信息的优化后,进行用户偏好分析的数据源更加准确,有利于得到更加有效的用户偏好数据。

[0066] 实施例六

[0067] 上述几个实施例所述的方法在得到用户的偏好数据后,信息推荐步骤的具体实现方法可以采用现有技术。现有技术通常是将内容信息独立和分别地进行推荐,这种推荐效果并不好。为此,本申请的实施例提供了一种协同推荐的方法,该方法通过对大量的用户内容信息分析得到内容信息的群体偏好数据,然后根据用户偏好数据和群体偏好数据进行内容信息推荐。本实施例除将实施例一的步骤S104变换为下述内容外,其他步骤相同。参见附图6,本实施例给出的内容信息推荐的步骤包括:

[0068] 步骤A':选取预设数量的其他用户组成参考用户集,计算所述用户与参考用户集内各用户的相似度,将相似度超过预设相似度阈值的用户组成相似用户集;

[0069] 根据用户的历史行为分析得到用户偏好,进而依据该用户偏好向用户推荐内容信息,固然迎合了用户的求购愿望,实现了提高交易成功率的目的,但是,实际上每个用户的用户偏好都不会局限在以往的用户偏好上,他的偏好是不断扩展变化的,而这种扩展变化往往受限于他自身的能力和获得信息的渠道,导致不能得到较好的扩展,此种情况下,与该用户具有相似或相同用户偏好的其他用户的用户偏好数据则值得借鉴,因为可以推知他们除具有上述相同的用户偏好外,还可能在其他方面也具有相同的偏好;基于这种分析,则需要选择一定数量的用户组成参考用户集,计算参考用户集中的用户与该用户的相似度,对相似度超过预设阈值的用户则构成该用户的相似用户集,相似用户集内的用户的偏好数据对该用户具有较高参考价值;需要说明的是计算用户之间的相似度可以采用多种方法,比如聚类算法,该方法在现有技术中已有大量应用,此处不再赘述。

[0070] 步骤B':分析相似用户集内各用户的用户行为得到该相似用户集的群体偏好数据,所述群体偏好数据为用户行为指向的内容信息的频率分布;

[0071] 通过步骤A'获得相似用户集后,分析相似用户集的各用户的用户行为得到群体偏好数据,具体的分析方法和前述实施例的方式相同,群体偏好数据是与单个用户的用户偏好数据“同质”的概念,只是数据量的不同在数据结果上存在差异,均表示用户行为指向的内容信息的频率分布情况。

[0072] 步骤C':根据用户偏好数据和所述群体偏好数据向用户推荐内容信息。

[0073] 本实施例通过计算相似度,找出与该用户具有相似特征的其他用户,并将其他用户的用户行为数据作为对该用户进行内容信息推荐的参考,较好地扩展了向用户推荐的内容信息,增强了用户体验,进一步拓宽了电子商务信息交易的领域和范围。

[0074] 本实施例在计算得到用户的相似用户集时,对“其他用户”的选取并没有作特别限定,实际上,由于地理环境的差异、各地风土人情的不同,同一个地域范围内用户的用户偏好数据可能更具有参考价值,因此,本实施例优选按照下面的方式选择参考用户集。

[0075] 判断所述用户IP地址对应的地域与该用户注册的地域是否一致,如果是,则选取所述用户注册地域地的其他用户组成参考用户集;如果否,则选取所述用户IP地址对应的地域地的其他用户组成参考用户集。

[0076] 通过以用户登录地的用户数据作为分析群体偏好的基础,不仅由于同一地域的用

户行为数据相比其他地方的用户行为数据的可参考性更大,使得分析得到的用户偏好结果更准确,据此进行推荐的推荐效果更好,极大提高了电子商务信息的交易成功率,而且由于采用就近原则,极大方便了用户在电子商务信息交易完成后尽快获得相应的产品或服务。

[0077] 实施例七

[0078] 在实施例六中已经提到得到用户的偏好数据后,信息推荐步骤的具体实现方法可以采用现有技术,也可以采用实施例六所述的协同推荐方法,协同推荐方法将待推荐用户的自身偏好数据与一定范围内的群体偏好数据结合起来,扩展了向待推荐用户推荐的内容信息,这是一种“横向”扩展,实际上,除横向扩展方式,本申请还提供一种纵向扩展方式的实施例,即将用户偏好数据中的多个用户偏好与预设规则集进行匹配,将匹配成功的某个或某些偏好作为向用户推荐内容信息的依据,这种方法称为联合推荐方法。本实施例除将实施例一的步骤S104变换为下述内容外,其他步骤相同。参见附图7,本实施例给出的内容信息推荐的步骤包括:

[0079] 步骤A":选取预设数量的用户组成规则创建用户集,分析所述规则创建用户集内各用户的用户行为得到各用户的用户偏好数据,根据各用户的用户偏好数据训练得到规则集,所述规则集为所述规则集为用以预测内容信息推荐的各类用户偏好的组合规则集;

[0080] 通过分析用户的历史行为得到的用户偏好数据,大多数情况下这种用户偏好是复合偏好,也就是说该用户偏好数据中包含了该用户在各个层面、各个领域、各的时段的多个偏好,比如产品偏好、供应商偏好、产品词偏好等,向用户进行内容信息推荐时尽管可以按照这些所有的偏好向用户推荐,但是,这种“广撒网”的方式在某些情况下并不能达到提高交易成功率的目的,反而容易导致用户在众多的内容信息中无所选择,因此,采用少而精、高命中率的方式向用户推荐内容信息将是一种可取的方式;为了从用户的众多偏好中找出“有效”的用户偏好,需要首先选取一定数量的用户组成规则创建用户集并根据这些用户的用户偏好建立规则集,进而利用该规则集作为选取用户偏好的依据;规则创建用户集内的各用户偏好数据的获取可以采用现有技术的方式得到,也可以按照本申请所提到方法对用户行为数据优化后分析得到;得到各用户的用户偏好数据后,将用户的各个偏好与该偏好对应的频率组成该用户的矢量特征组,然后对所有的矢量特征组进行训练得到规则集,具体的训练方式现有技术已公开,这里不再赘述。

[0081] 步骤B":将待推荐用户的用户偏好数据与所述规则集进行匹配,根据匹配成功的偏好数据向待推荐用户推荐内容信息。

[0082] 获得到规则集后,将待推荐用户的用户偏好与规则集中的规则进行匹配,如果匹配成功则将根据相应的用户偏好向该用户推荐内容信息。

[0083] 为了更加清楚的说明上述过程,这里举例阐释:假设选取的规则创建用户集为100个,每个用户的偏好包括产品、供应商、产品词、求购、资讯等五个偏好,通过分析100个用户的历史行为数据得到了各个用户的偏好数据,偏好数据包括内容信息和该内容信息的频率,为了便于比较将频率进行归一化处理,该归一化的基准可以是10次;通过按照现有技术的方法训练100个用户的用户偏好数据得到三条规则,现在需要向A1、B2、C3三个用户进行推荐,推荐时将A1的用户偏好数据与规则集内的规则进行匹配,然后按照匹配成功的用户偏好推荐内容信息。上述过程请参见如下表格:

[0084]

用户名称		A1	B2	C3	规则集
用户偏好类别					
产品偏好	归一化频率 偏好指向类目	0.6 数学	0 无	0.2 数学	规则 1: 如果产品偏好的归一化频率大于 0.5, 则推荐产品偏好对应内容信息;
供应商偏好	归一化频率 偏好指向类目	0.2 英语	0 无	0.1 数学	
产品词偏好	归一化频率 偏好指向类目	0.3 数学	0 无	1 语文	规则 2: 如果产品词偏好的归一化频率大于产品偏好与资讯偏好归一化频率之和的 1.2 倍, 则推荐产品词偏好对应内容信息;
求购偏好	归一化频率 偏好指向类目	0 无	1 语文	0 无	
资讯偏好	归一化频率 偏好指向类目	0 无	0.5 英语	0.5 英语	规则 3: 如果群体偏好的归一化频率大于资讯品号与求购偏好归一化频率之和的 2 倍, 则推荐群体偏好对应内容信息。
群体偏好	归一化频率 偏好指向类目	0 无	4 英语	0 无	
匹配规则集结果		符合规则 1	符合规则 3	符合规则 2	
推荐内容信息		产品偏好 数学	群体偏好 英语	产品词偏好 语文	

[0085] 值得注意的是上述表格中列出了群体偏好一栏,前面介绍了根据规则创建用户集的各用户的用户行为分析得到了用户偏好数据,实际上,也可以根据该集合中的用户偏好数据进一步得到这个群体的群体偏好,进而利用该群体偏好来创建关于群体偏好的规则。

[0086] 本实施例将用户的多种偏好结合起来,选择最优化的推荐方案进行内容信息推荐,避免了传统推荐方法根据所有的偏好毫无取舍的“全盘”推荐,增强了推荐效果,有效提高了交易成功率。

[0087] 实施例八

[0088] 上述七个实施例均是本申请的方法实施例,对应地,本申请还提供了装置实施例。参见附图8,本申请的实施例提供的一种基于用户行为的电子商务信息推荐装置800包括:

[0089] 统计模块801、判断模块802、偏好分析模块803和信息推荐模块804,其中:

[0090] 统计模块801,用于统计预设时期内用户行为类型和各行为类型对应的用户行为的频率;

[0091] 判断模块802,用于判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型;和/或,判断各用户行为的频率是否超过预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率;

[0092] 偏好分析模块803,用于根据上述处理后的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到用户偏好数据,所述用户偏好数据为各内容信息的频率分布;

[0093] 804信息推荐模块,用于利用用户偏好数据向用户推荐内容信息。

[0094] 本装置的工作过程是:统计模块801统计预设时期内用户行为类型和各行为类型对应的用户行为的频率,并将数据传递给判断模块802,由该模块判断用户行为类型数是否超过预设类型阈值,如果是,则保留小于等于预设类型阈值数的用户行为类型;和/或,判断各用户行为的频率是否超过该用户行为的预设频率阈值,如果是,则将预设频率阈值作为该用户行为的频率;然后由偏好分析模块803根据上述处理后的用户行为类型和用户行为

的频率分析用户行为指向的内容信息,得到用户偏好数据,并将该数据传递给信息推荐模块804,由该模块利用用户偏好数据进行内容信息推荐。

[0095] 本申请的技术方案在根据用户偏好进行信息推荐之前,通过判断用户行为类型数与用户行为频率是否符合预设值的方式对用户的历史行为进行优化,优化后的用户行为数据排除了恶意行为和作弊行为,在该用户行为数据基础上分析得到的用户偏好数据较为真实地反映了用户的偏好特点,进而使得根据该用户偏好数据进行信息推荐的推荐效果更好,有利于促进电子商务信息交易的正常发展。此外,由于推荐的商品信息能够符合用户的实际需求愿望,这样用户只需根据电子商务交易平台推荐的信息完成交易即可,而不必再反复查找其所需的商品信息,降低了反复查找商品信息给电子商务平台带来的访问压力。并且,由于本申请的优化处理过程去除了用户行为数据中的大量“伪数据”,减少了用于信息推荐分析的数据量,提高了信息推荐分析的速度,能够更加快速的获取到推荐信息。

[0096] 本实施例的装置800还可以包括时间影响消除模块805,用于在分析用户行为指向的内容信息前,对用户行为赋予时间权重值,所述时间权重值为越靠近预设时期结束时刻权重值越高,越靠近预设时期起始时刻权重值越低。该模块消除时间影响后有利于更加准确地分析用户行为数据。

[0097] 本实施例的装置800偏好分析模块803可以有多种组成机构,只要能实现本申请的发明目的即可,但是,本申请的实施例优选按照如下结构方式组成该模块,该模块包括:

[0098] 时段单元划分子模块8031、偏好分析子模块8032和过滤子模块8033,其中:

[0099] 时段单元划分子模块8031,用于根据用户行为发生的时间将处理后的用户行为类型和用户行为的频率在预设时期内划分成至少两个时段单元;

[0100] 偏好分析子模块8032,用于根据各时段单元内的用户行为类型和用户行为的频率分析用户行为指向的内容信息,得到该时段单元的用户偏好数据;

[0101] 过滤子模块8033,用于根据各时段单元的用户偏好数据过滤掉迁移性、突发性偏好数据以得到预设时期的用户偏好数据。

[0102] 本实施例的装置800还可以包括相似度计算模块806,用于在选取预设数量的其他用户组成参考用户集后,计算待推荐用户与参考用户集内各用户的相似度,将相似度超过预设相似度阈值的用户组成相似用户集;则:偏好分析模块803还用于分析相似用户集内各用户的用户行为得到该相似用户集的群体偏好数据;信息推荐模块804,还用于根据待推荐用户的用户偏好数据和所述群体偏好数据向该用户推荐内容信息。

[0103] 本实施例的装置800还包括规则集创建模块807,用于在选取预设数量的用户组成规则创建用户集,调用偏好分析模块分析所述规则创建用户集内各用户的用户行为得到各用户的用户偏好数据,根据各用户的用户偏好数据训练得到规则集,所述规则集为用以预测内容信息推荐各类用户偏好的组合规则集,则:偏好分析模块还用于分析规则创建用户集内各用户的用户行为得到各用户的用户偏好数据,所述用户偏好数据为用户行为指向的内容信息的频率分布;信息推荐模块,还用于将待推荐用户的用户偏好数据与所述规则集进行匹配,根据匹配成功的用户偏好数据向待推荐用户推荐内容信息。包含规则集创建单元807的装置将用户的多种偏好结合起来,选择最优化的推荐方案进行内容信息推荐,避免了传统推荐方法根据所有的偏好毫无取舍的“全盘”推荐,增强了推荐效果,有效提高了交易成功率。

[0104] 以上所述仅为本申请的较佳实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在发明的保护范围之内。

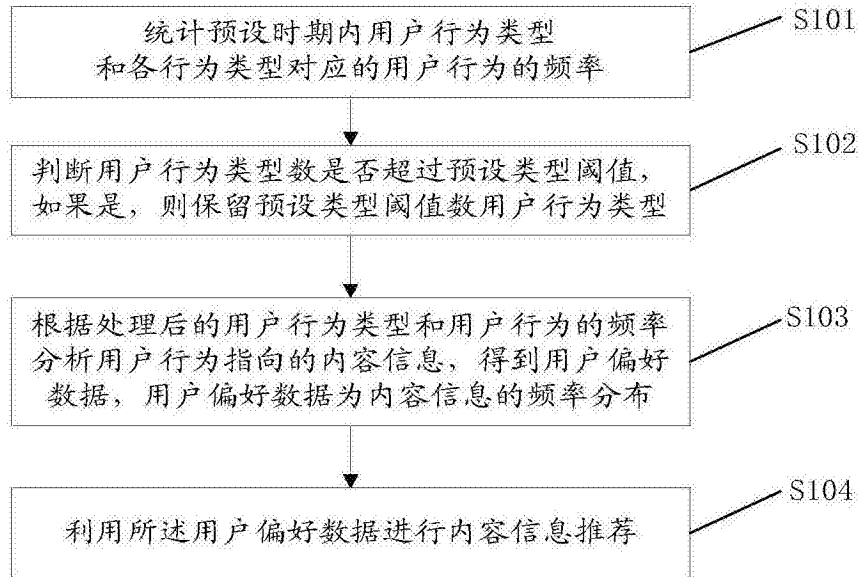


图1

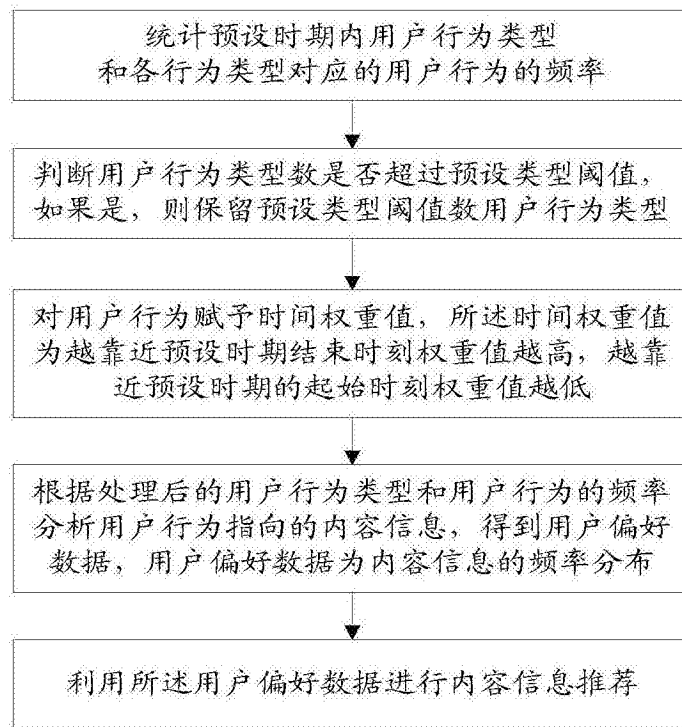


图2

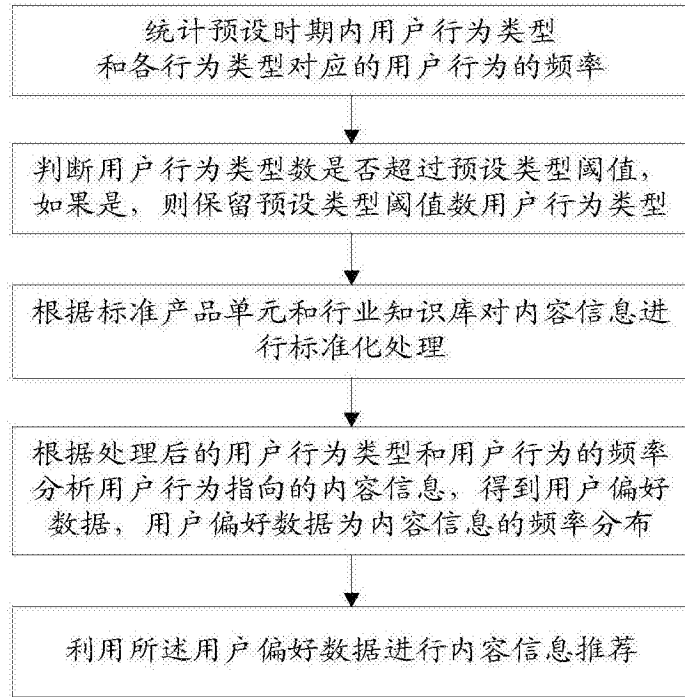


图3

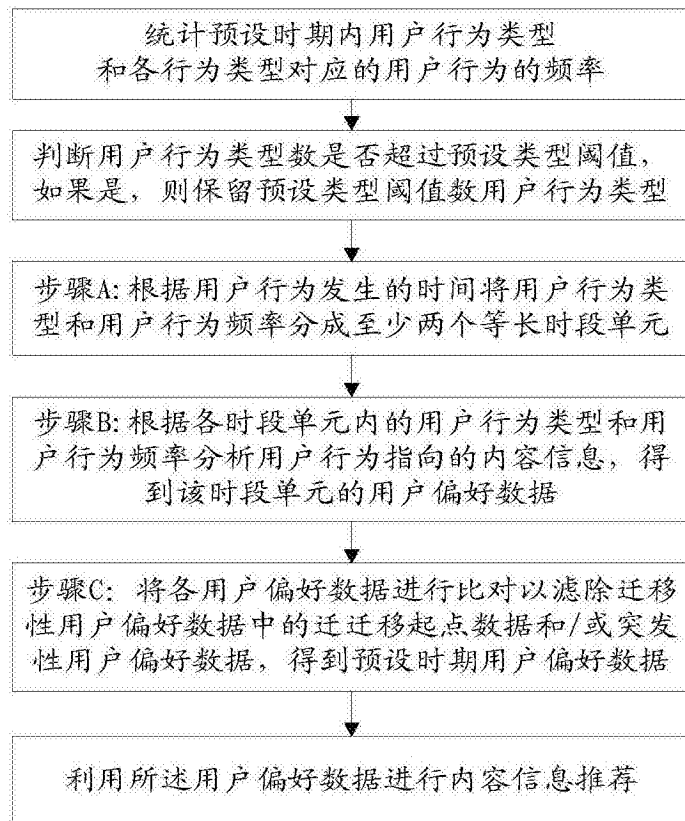


图4

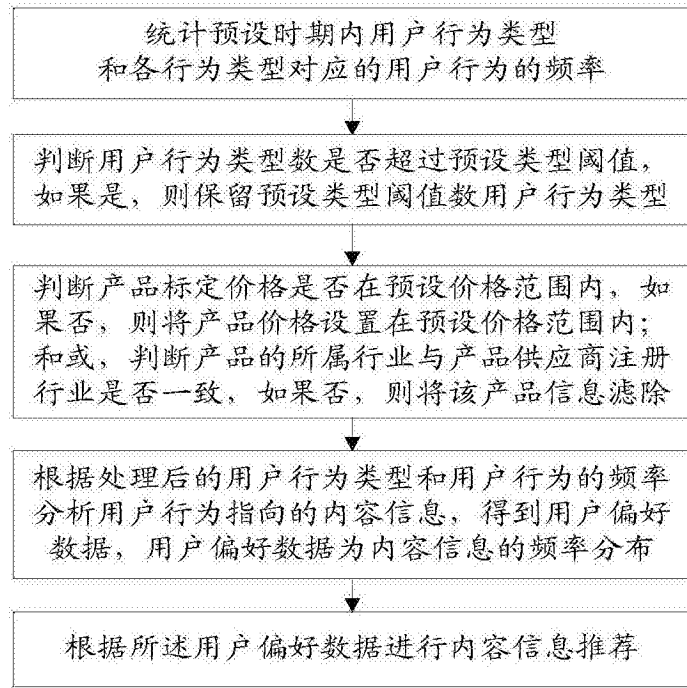


图5



图6

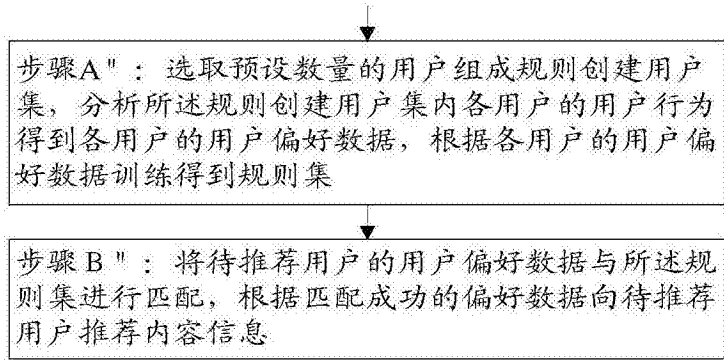


图7

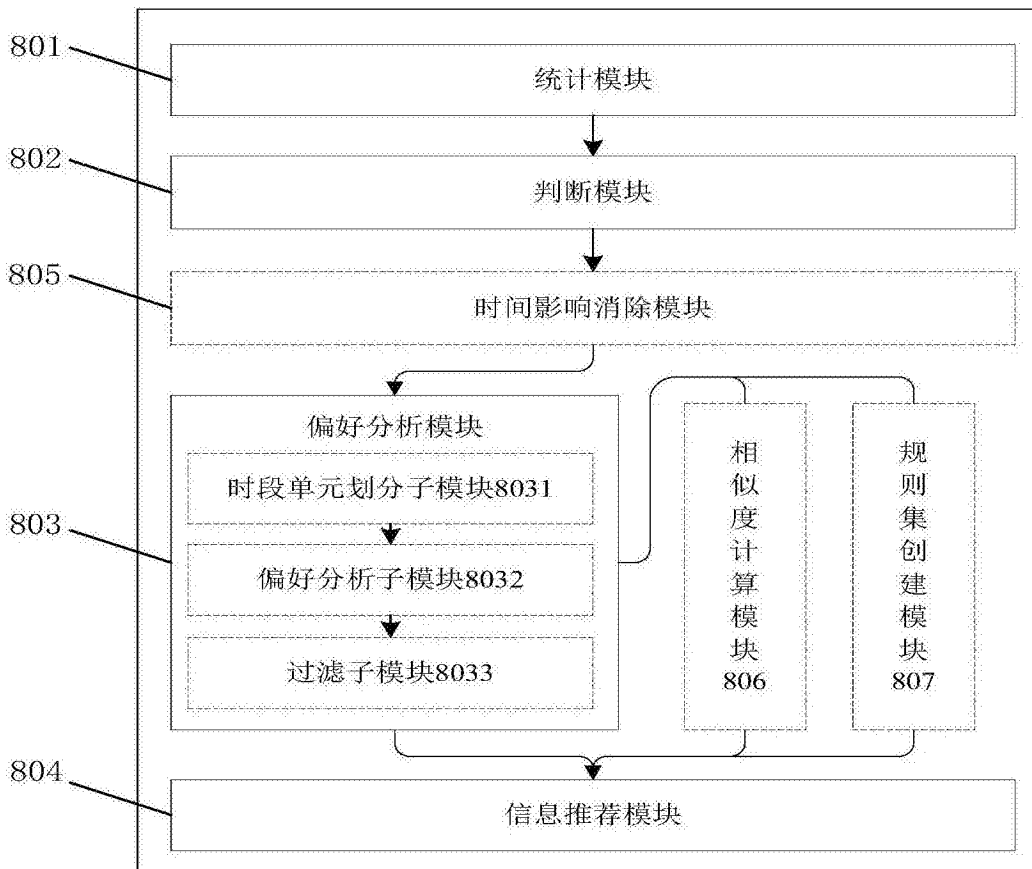


图8