



(12) 发明专利申请

(10) 申请公布号 CN 103607496 A

(43) 申请公布日 2014. 02. 26

(21) 申请号 201310573351. 7

(22) 申请日 2013. 11. 15

(71) 申请人 中国科学院深圳先进技术研究院
地址 518055 广东省深圳市南山区西丽大学
城学苑大道 1068 号

(72) 发明人 李翔宇 张潇

(74) 专利代理机构 深圳中一专利商标事务所
44237

代理人 张全文

(51) Int. Cl.

H04M 1/2745(2006. 01)

G06F 17/30(2006. 01)

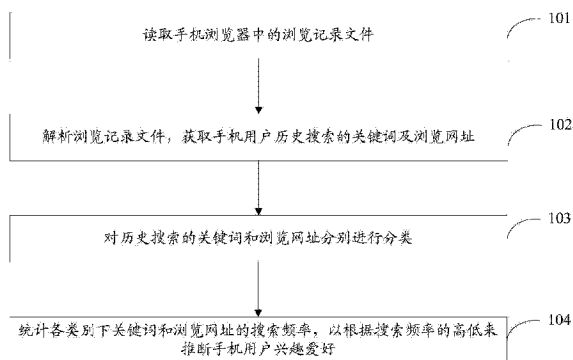
权利要求书1页 说明书5页 附图2页

(54) 发明名称

一种推断手机用户兴趣爱好方法、装置及手机终端

(57) 摘要

本发明适用于通讯技术领域,提供了一种推断手机用户兴趣爱好方法、装置及手机终端,所述方法包括:读取手机浏览器中的浏览记录文件;解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址;对历史搜索的关键词和浏览网址分别进行分类;统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。本发明通过读取浏览记录文件、获取关键词和浏览网址,并统计各类别下关键词和浏览网址的搜索频率,可有效的根据频率的高低来推断手机用户兴趣爱好。



1. 一种推断手机用户兴趣爱好的方法,其特征在于,所述方法包括:
读取手机浏览器中的浏览记录文件;
解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址;
对历史搜索的关键词和浏览网址分别进行分类;
统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。
2. 如权利要求1所述的方法,其特征在于,所述对历史搜索的关键词和浏览网址分别进行分类的步骤包括:
对历史搜索的关键词按照语义分类;
对历史搜索的浏览网址按照域名由高到低层次分类。
3. 如权利要求2所述的方法,其特征在于,所述方法还包括:
采用数组的方式存储各类别下的关键词及对应的关键词的输入搜索频率,同一类别下的不同关键词通过建立的数组下标索引来标记区分;
采用B+树的方式存储按域名由高到低层次分类的浏览网址及对应的浏览网址的搜索访问频率。
4. 如权利要求1或2或3所述的方法,其特征在于,所述方法还包括:
按照手机用户兴趣爱好向手机用户推荐关键词、网站或应用。
5. 如权利要求1所述的方法,其特征在于,所述浏览记录文件包括日志文件和缓存文件。
6. 一种推断手机用户兴趣爱好的装置,其特征在于,所述装置包括:
读取单元,用于读取手机浏览器中的浏览记录文件;
获取单元,用于解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址;
分类单元,用于对历史搜索的关键词和浏览网址分别进行分类;
统计推断单元,用于统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。
7. 如权利要求6所述的装置,其特征在于,
所述分类单元,具体用于对历史搜索的关键词按照语义分类和对历史搜索的浏览网址按照域名由高到低层次分类。
8. 如权利要求7所述的装置,其特征在于,所述装置还包括:
数组存储单元,用于采用数组的方式存储各类别下的关键词及对应的关键词的输入搜索频率,同一类别下的不同关键词通过建立的数组下标索引来标记区分;
B+树存储单元,用于采用B+树的方式存储按域名由高到低层次分类的浏览网址及对应的浏览网址的搜索访问频率。
9. 如权利要求6或7或8所述的装置,其特征在于,所述装置还包括:
推荐单元,用于按照手机用户兴趣爱好向手机用户推荐关键词、网站或应用。
10. 一种手机终端,其特征在于,所述手机终端包括权利要求6所述的装置。

一种推断手机用户兴趣爱好方法、装置及手机终端

技术领域

[0001] 本发明属于通讯技术领域,尤其涉及一种推断手机用户兴趣爱好方法、装置及手机终端。

背景技术

[0002] 随着手机的普及,手机用户不断增加,手机网民也在不断增加。据有关数据统计,2011年国内手机用户总量已达到9.3亿,且手机上网用户已超过3.9亿。此外,据DCCI互联网数据中心预测,到2013年中国手机网民将达7.20亿,手机网民将超越电脑网民。手机浏览器作为网民通过手机浏览网页的工具,具有非常好的发展前景。

[0003] 现有技术中,普通的手机终端不具备推断手机用户兴趣爱好的功能。

发明内容

[0004] 本发明实施例的目的在于提供一种推断手机用户兴趣爱好的方法,旨在解决普通的手机终端不具备推断手机用户兴趣爱好的功能的问题。

[0005] 本发明实施例是这样实现的,一种推断手机用户兴趣爱好的方法,所述方法包括:

[0006] 读取手机浏览器中的浏览记录文件;

[0007] 解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址;

[0008] 对历史搜索的关键词和浏览网址分别进行分类;

[0009] 统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。

[0010] 本发明实施例还提供了一种推断手机用户兴趣爱好的装置,所述装置包括:

[0011] 读取单元,用于读取手机浏览器中的浏览记录文件;

[0012] 获取单元,用于解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址;

[0013] 分类单元,用于对历史搜索的关键词和浏览网址分别进行分类;

[0014] 统计推断单元,用于统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。

[0015] 本发明实施例还提供了一种手机终端,所述手机终端包括上述的装置。

[0016] 本发明实施例与现有技术相比,有益效果在于:通过读取浏览记录文件、获取关键词和浏览网址,并统计各类别下关键词和浏览网址的搜索频率,可有效的根据频率的高低来推断手机用户兴趣爱好。

附图说明

[0017] 图1是本发明实施例提供的推断手机用户兴趣爱好的方法的流程图;

[0018] 图2是本发明实施例提供的推断手机用户兴趣爱好的装置的第一逻辑示意图;

[0019] 图 3 是本发明实施例提供的推断手机用户兴趣爱好的装置的第二逻辑示意图；

[0020] 图 4 是本发明实施例提供的推断手机用户兴趣爱好的装置的第三逻辑示意图。

具体实施方式

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

[0022] 本发明实施例提供的实施方案如下：

[0023] 为了方便理解本发明实施例，首先在此介绍本发明实施例描述中会引入的几个要素：

[0024] 域名(Domain Name)：

[0025] 域名，是由一串用点分隔的名字组成的 Internet 上某一台计算机或计算机组的名称，用于在数据传输时标识计算机的电子方位(有时也指地理位置，地理上的域名，指代有行政自主权的一个地方区域)。域名是一个 IP 地址上有“面具”。一个域名的目的是便于记忆和沟通的一组服务器的地址(网站，电子邮件，FTP 等)。域名作为力所能及难忘的互联网参与者的名称，如电脑，手机终端、网络和服务。

[0026] 一个完整的域名由二个或二个以上部分组成，各部分之间用英文的句号“.”来分隔，例如下列域名：yahoo.com, yahoo.ca.us, yahoo.co.uk。其中第一个域名由二部分组成，第二个域名和第三个域名由三部分组成。在一个完整的域名中，最后一个“.”的右边部分称为顶级域名或一级域名(TLD)，在上面的域名例子中，com、us 和 uk 是顶级域名。最后一个“.”的左边部分称为二级域名(SLD)，例如，域名 yahoo.com 中 yahoo 是二级域名，域名 yahoo.ca.us 中 ca 是二级域名，而域名 yahoo.co.uk 中 co 是二级域名。二级域名的左边部分称为三级域名，三级域名的左边部分称为四级域名，以此类推。例如，域名 yahoo.ca.us 和 yahoo.co.uk 中 yahoo 是三级域名。

[0027] B+ 树的定义和特点：

[0028] 一、一棵 m 阶的 B+ 树定义为：

[0029] (1) 每个节点最多可以有 m 个元素；

[0030] (2) 除了根节点外，每个节点最少有 $(m/2)$ 个元素；

[0031] (3) 如果根节点不是叶节点，那么它最少有 2 个孩子节点；

[0032] (4) 所有的叶子节点都在同一层；

[0033] (5) 一个有 k 个孩子节点的非叶子节点有 $(k-1)$ 个元素，按升序排列；

[0034] (6) 某个元素的左子树中的元素都比它小，右子树的元素都大于或等于它；

[0035] (7) 非叶子节点只存放关键字和指向下一个孩子节点的索引，记录只存放在叶子节点中；

[0036] (8) 相邻的叶子节点之间用指针相连。

[0037] 二、B+ 树具有的特点是：

[0038] (1) 所有关键字都出现在叶子结点的链表中(稠密索引)，且链表中的关键字恰好是有序的；

[0039] (2) 不可能在非叶子结点命中；

[0040] (3) 非叶子结点相当于是叶子结点的索引(稀疏索引),叶子结点相当于是存储(关键字)数据的数据层;

[0041] (4) 更适合文件索引系统;

[0042] 请参阅图 1,本发明实施例提供一种推断手机用户兴趣爱好的方法,所述方法包括:

[0043] 101、读取手机浏览器中的浏览记录文件。

[0044] 手机浏览器中的浏览记录文件包括日志文件和缓存文件。

[0045] 手机用户通过手机自带的手机浏览器或者安装的其他手机浏览器软件上网,所有的浏览记录都记录在对应浏览器的日志文件和缓存文件中。

[0046] 对于不同的浏览器,它的日志文件和缓存文件存储在手机上的位置不同,读取时可以通过文件的后缀名来区分,因为不同的浏览器的日志文件和缓存文件有不同的后缀名。

[0047] 102、解析浏览记录文件,获取手机用户历史搜索的关键词及浏览网址。

[0048] 通过解析日志文件和缓存文件,就可以获取手机用户的浏览记录,该浏览记录包括用户之前输入搜索的关键词、之前的浏览网址。这些关键词和浏览网址反映了用户关注的领域以及兴趣爱好。

[0049] 通过解析浏览记录文件的方式来获取手机用户历史搜索的关键词和浏览网址,简单且操作方便。

[0050] 103、对历史搜索的关键词和浏览网址分别进行分类。

[0051] 本实施例中,所述对历史搜索的关键词和浏览网址分别进行分类的步骤包括:

[0052] 对历史搜索的关键词按照语义分类;

[0053] 对历史搜索的浏览网址按照域名由高到低层次分类。

[0054] 本实施例中,优选的,所述方法还包括:

[0055] 采用数组的方式存储各类别下的关键词及对应的关键词的输入搜索频率,同一类别下的不同关键词通过建立的数组下标索引来标记区分;

[0056] 采用 B+ 树的方式存储按域名由高到低层次分类的浏览网址及对应的浏览网址的搜索访问频率。

[0057] 本实施例中,通过数组的方式存储关键词及其输入搜索的频率,还可通过数组的下标建立索引;通过 B+ 树的方式存储浏览网址及搜索访问频率,查询时可以按照链表查询或者树查询,这两种方式方便建立查询索引,有利于排序和查找,执行效率高。

[0058] 本实施例中,将用户所用的关键词按照语义相近程度分类,每一类中取频率最高的关键词,当用户再次打开浏览器准备输入关键词时,可以将这些关键词推荐给用户。

[0059] 本实施例中,103 之前还包括以下步骤:

[0060] 预先设置关键词的类别。

[0061] 优选的,可提供一个设置界面,让用户根据自己的需要设置常用的类别。在实际应用中,可只按照用户设置的类别来进行分类,从而通过与用户的互动,可更好的推断用户的兴趣爱好,或者说由用户自己提供自己的兴趣爱好,更加直接方便。对用户没有设置的类别不予考虑。

[0062] 对关键词进行分类时,可提前设定好类别,例如按照语义可以分为娱乐,学习,办

公, 休闲等等, 首先将所有的关键词归类, 将关键词划分到特定的大类中, 同义关键词划归到一类。然后, 对特定的大类中的关键词按照频率由高到低排列。再将大类中的频率最高的关键词取出再按照频率高低排列, 以此顺序排列后确定用户的兴趣爱好。关键词的使用频率高低代表了用户具体的兴趣爱好。

[0063] 对用户访问的网站的访问次数进行统计, 使用 B+ 树的数据结构, 将网址按照域名由高到低的层次分类统计, 比如用户访问百度, 百度包含百度百科, 百度图片, 百度新闻等, 而百度百科具体的内容对应的网站属于最高域名的网址, 其包含的内容才是用户的搜索的最终结果。统计用户浏览网站的频率, 将用户浏览指定网站的最高层次的域名的网站所属的类别作为用户兴趣爱好的类别, 将访问域名最高层次的网站按照频率高低排列, 依据排列结果推断用户的兴趣爱好。

[0064] 104、统计各类别下关键词和浏览网址的搜索频率, 以根据搜索频率的高低来推断手机用户兴趣爱好。

[0065] 本实施例中, 步骤 104 具体包括以下步骤:

[0066] 统计各类别下关键词的搜索频率;

[0067] 将各类别中搜索频率最高的关键词按照频率由高到低排序;

[0068] 统计各类别下浏览网址的搜索频率;

[0069] 将各类别中最高域名的网址按照频率由高到低排序;

[0070] 根据两个排序来推断手机用户的兴趣爱好。

[0071] 本实施例中, 将用户使用的关键词和浏览网址划分到相应的类别中, 再将相应类别中的关键词和浏览网址按照频率由高到低排列统计出来, 从而可得出用户最关注的类别和最不关注的类别, 以此推断出用户的兴趣爱好。

[0072] 本实施例中, 优选的, 所述方法还包括:

[0073] 按照手机用户兴趣爱好向手机用户推荐关键词、网站或应用。

[0074] 按照手机用户兴趣爱好, 可以向其推荐关键词、相关的网站或者应用。可为手机浏览器开发商提供便利, 采用本实施例的方法可方便向用户推荐常用的搜索关键词、还可在浏览的网页里给用户推荐与其兴趣爱好相关的网站, 不但增加浏览器的功能、增强了手机操作系统及软件的体验, 方便了手机用户, 浏览器开发商还可以此赚取广告的推荐费用, 带来经济效益。

[0075] 本实施例中, 优选的, 所述方法还包括:

[0076] 向手机用户推荐各类别中搜索频率最高的关键词;

[0077] 向手机用户推荐各类别中搜索频率最高的最高域名的网址。

[0078] 可给经常使用手机浏览器浏览输入某固定关键词或是访问某固定网站的手机用户带来便利。

[0079] 请参阅图 2, 本发明实施例还提供了一种推断手机用户兴趣爱好的装置, 所述装置包括:

[0080] 读取单元 201, 用于读取手机浏览器中的浏览记录文件;

[0081] 获取单元 202, 用于解析浏览记录文件, 获取手机用户历史搜索的关键词及浏览网址;

[0082] 分类单元 203, 用于对历史搜索的关键词和浏览网址分别进行分类;

[0083] 优选的,分类单元 203 还包括类别模块,用于预先设置关键词的类别。

[0084] 优选的,可提供一个设置界面,让用户根据自己的需要设置常用的类别。在实际应用中,可只按照用户设置的类别来进行分类,从而通过与用户的互动,可更好的推断用户的兴趣爱好,或者说由用户自己提供自己的兴趣爱好,更加直接方便。对用户没有设置的类别不予考虑。

[0085] 统计推断单元 204,用于统计各类别下关键词和浏览网址的搜索频率,以根据搜索频率的高低来推断手机用户兴趣爱好。

[0086] 优选的,所述分类单元,具体用于对历史搜索的关键词按照语义分类和对历史搜索的浏览网址按照域名由高到低层次分类。

[0087] 请参阅图 3,本实施例中,优选的,所述装置还包括:

[0088] 数组存储单元 301,用于采用数组的方式存储各类别下的关键词及对应的关键词的输入搜索频率,同一类别下的不同关键词通过建立的数组下标索引来标记区分;

[0089] B+ 树存储单元 302,用于采用 B+ 树的方式存储按域名由高到低层次分类的浏览网址及对应的浏览网址的搜索访问频率。

[0090] 请参阅图 4,所述装置还包括:

[0091] 推荐单元 401,用于按照手机用户兴趣爱好向手机用户推荐关键词、网站或应用。

[0092] 装置中的细节方案已在方法中描述,在此不再赘述。

[0093] 本发明实施例还提供一种手机终端,所述手机终端包括上述的装置。

[0094] 本发明的推断手机用户兴趣爱好的方法、装置及手机终端,通过读取浏览记录文件、获取关键词和浏览网址,并统计各类别下关键词和浏览网址的搜索频率,可有效的根据频率的高低来推断手机用户兴趣爱好。

[0095] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

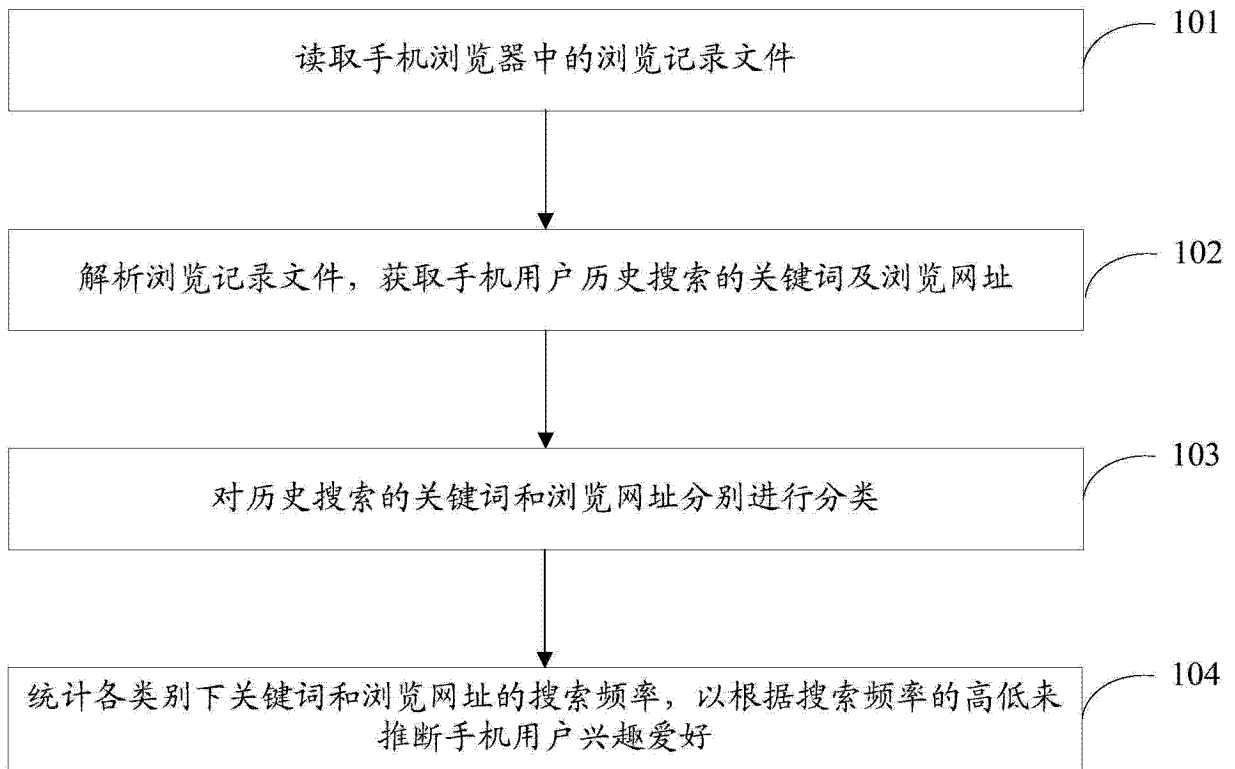


图 1

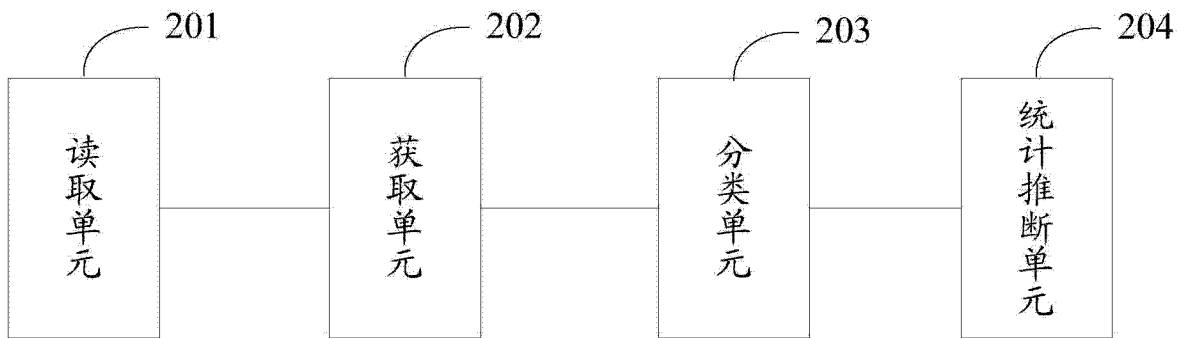


图 2

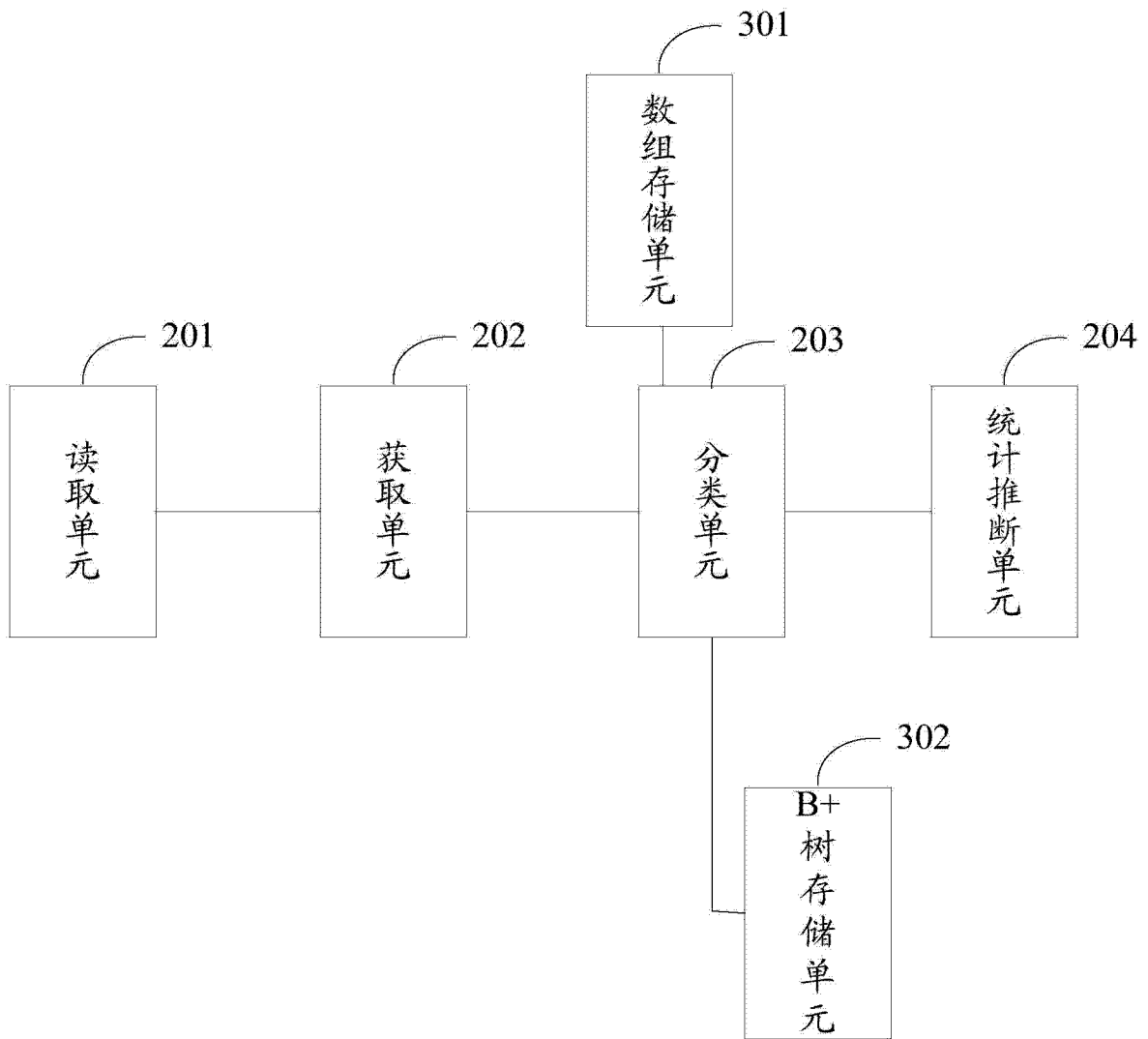


图 3

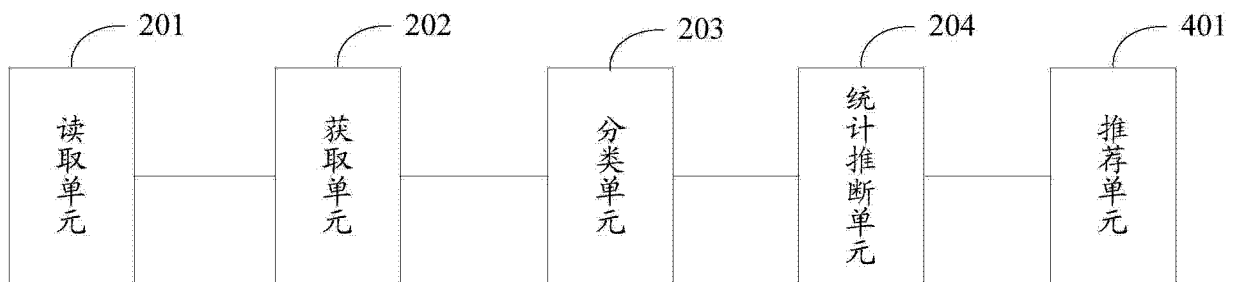


图 4