(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2019 年 12 月 5 日 (05.12.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/227650 A1

(51) 国际专利分类号:
G06F 11/34 (2006.01)

(21) 国际申请号: PCT/CN2018/097448

(22) 国际申请日: 2018 年 7 月 27 日 (27.07.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810554442.9 2018年6月1日 (01.06.2018) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 朱坤 (ZHU, Kun); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一专利商标事务所 (SHENZHEN ZHONGYI PATENT AND TRADEMARK OFFICE); 中国广东省深圳市福田区深南中路1014号老特区报社四楼（5号信箱）, Guangdong 518028 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,

(54) **Title:** EVENT TRACKING DATA ANALYSIS METHOD, TERMINAL APPARATUS AND COMPUTER READABLE STORAGE MEDIUM

(54) **发明名称**: 一种埋点数据的分析方法、终端设备及计算机可读存储介质

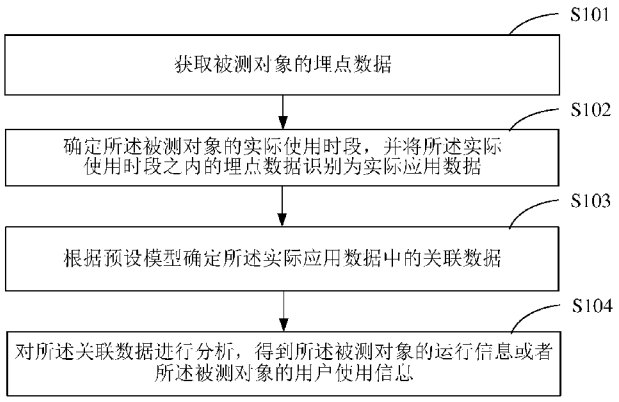

图 1

S101 Acquire event tracking data of a detected object
S102 Determine an actual time of use of the detected object and identify event tracking data within the actual time of use as actual application data
S103 Determine associated data in the actual application data according to a preset model
S104 Analyze the associated data to obtain operation information of the detected object or usage information of a user of the detected object

(57) **Abstract:** An event tracking data analysis method and a device, comprising: acquiring event tracking data of a detected object (S101), the detected object comprising an application or a web page; determining an actual time of use of the detected object, and identifying event tracking data within the actual time of use as actual application data (S102); determining associated data in the actual application data according to a preset model (S103); and analyzing the associated data to obtain operation information of the detected object or usage information of a user of the detected object (S104). The associated data is analyzed to obtain circumstances of user behavior and a software operation state within the entire actual time of use, thereby realizing complete and comprehensive data compilation and analysis, and increasing the coherence and trackability of user behavior analysis.

WO 2019/227650 A1

[见续页]

RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种埋点数据的分析方法及装置，包括：获取被测对象的埋点数据(S101)；所述被测对象包括应用程序或网页；确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据(S102)；根据预设模型确定所述实际应用数据中的关联数据(S103)；对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息(S104)。通过分析关联数据得到整个实际使用时间之内的用户行为情况和软件的运行状态情况，实现完整、全面的数据统计与分析，提高了用户行为分析的连贯性和可追踪性。

一种埋点数据的分析方法、终端设备及计算机可读存储介质

本申请申明享有2018年4月9日递交的申请号为201810554442.9、名称为“一种埋点数据的分析方法及装置”中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请属于数据处理技术领域，尤其涉及一种埋点数据的分析方法及装置。

背景技术

在应用程序（Application，App）或者网页中进行埋点便可以获取到用户在使用过程中的操作信息，通过在App代码中加入与信息搜集相关的代码，在适当的时机发送给服务器，服务器通过数据建模等方法确定用户的具体动作。例如，如果开发者想要知道用户使用某个软件的情况，那么在软件的某个按钮的监听函数中加上一段代码，用一个变量记下用户点击该按钮的次数，然后通过合理的时机将这个变量的值传给该软件后台的服务器，服务器对这些变量的值进行分析，便可确定出用户打开软件之后的具体动作。

但是，现有技术中在获取到用户的应用数据之后，只是根据埋点处获取到的操作信息单独分析某一点处的用户数据，尤其是当App或者网页的功能较复杂、用户数量技术较大的情况下，不能对整个使用过程进行全面的监控，造成数据分析的局限性。

技术问题

本申请实施例提供了一种埋点数据的分析方法及装置，以解决现有技术中单独分析某一点处的用户数据，造成数据分析的局限性，而不能对整个使用过程进行全面监控的问题。

技术解决方案

第一方面，提供了一种埋点数据的分析方法，包括：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

第二方面，提供了一种终端设备，包括：

埋点数据获取单元，用于获取被测对象的埋点数据；所述被测对象包括应用程序或网页。

实际数据确定单元，用于确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据。

关联数据确定单元，用于根据预设模型确定所述实际应用数据中的关联数据。

关联数据分析单元，用于对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

第三方面，提供了一种服务器，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

第四方面，提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

有益效果

实施本申请实施例提供的一种埋点数据的分析方法及装置具有以下有益效果：通过获取埋点数据确定被测对象的实际使用时段，以及该实际使用时段之内的关联数据，通过分

析关联数据得到整个实际使用时间之内的用户行为情况和软件的运行状态情况，实现完整、全面的数据统计与分析，提高了用户行为分析的连贯性，以提高网页或者应用程序的可追踪性。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请实施例一提供的埋点数据的分析方法的流程图；

图 2 是本申请实施例二提供的埋点数据的分析方法的流程图；

图 3 是本申请实施例二提供的实际使用时段的示意图；

图 4 是本申请实施例三提供的埋点数据的分析方法的流程图；

图 5 是本申请实施例四提供的终端设备的示意图；

图 6 是本申请实施例五提供的终端设备的示意图。

本发明的实施方式

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，此处所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。

本申请实施例通过埋点数据得到关联数据，以对关联数据进行分析，得到整个时间段之内、某一操作过程中的用户行为情况和软件的运行状态情况，解决了根据单一埋点处获取到的操作信息单独分析某一点处的用户数据，造成数据分析的局限性，而不能对整个使用过程进行全面监控的问题。

实施例 1

在本申请实施例中，流程的执行主体为计算机、服务器等终端设备，但并不限于此，该终端设备具有分析埋点数据的功能。图 1 示出了本申请第一实施例提供的终端设备的控制方法的实现流程图，详述如下：

在 S101 中，获取被测对象的埋点数据；所述被测对象包括应用程序或网页。

埋点是分析被测对象时的一种数据采集方法，通过在被测对象的程序中进行埋点可以

获取到被测对象的运行数据，再对这些数据进行分析便可以得到被测对象的运行情况，或者用户使用该被测对象时的用户习惯等。

示例性地，在实际应用中，在应用程序的界面初始化时，同时初始化第三方数据分析服务商的软件开发工具包（Software Development Kit，SDK），然后在某个事件发生时就调用 SDK 里面相应的数据发送接口发送数据。例如，若需要统计应用程序里面某个按钮的点击次数，则在这个按钮被点击时，在按钮对应的点击鼠标 OnClick 函数里面调用 SDK 提供的数据发送接口来发送数据。

进一步的，在实际获取被测对象的埋点数据时，可以是设定一个或者多个固定的时间来获取埋点数据；也可以设定一个获取周期，周期性的获取埋点数据；还可以检测当前时刻的网络状态是否拥堵，根据所生成的埋点数据量的多少，选择网络状态良好的时候获取埋点数据，此处不做限定。

需要说明的是，本实施例中的被测对象可以是计算机或者终端设备中的应用程序，也可以是其中的网页，此处不做限定。

在 S102 中，确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据。

由于被测对象在实际运行过程中，可能一直在生成埋点数据。若埋点数据的数据量较大，且冗余数据较多时，往往会对分析过程造成干扰，而影响埋点数据的分析结果。例如，在某个被测对象被开启并未进行操作时，便可能生成埋点数据，这些埋点数据即为冗余数据。可选的，可以将这些冗余数据直接删除，以保证埋点数据的精简性，提高数据传输和处理的效率。

在本实施例中，通过根据埋点数据的属性，确定被测对象的实际使用时段，以将使用时段之内的埋点数据识别为实际应用数据，保证数据分析的有效性。

在实际应用中，每个软件或者网页在实际使用过程中，都是从打开到转入后台、锁屏或者关闭，在这个过程中所有的操作动作都是可以通过埋点数据体现的。例如，打开或者关闭一个应用程序时，其埋点数据中就会体现出应用程序的启动时间或者关闭时间。

每个埋点数据中都有时间标识，这个时间标识用来确定生成该埋点数据的时间。通过识别埋点数据的变化特点，确定被测对象的实际使用时间，并根据该实际使用时间和埋点数据的时间标识，截取在该实际使用时间之内的埋点数据，这些截取到的数据，便是实际应用数据。

在 S103 中，根据预设模型确定所述实际应用数据中的关联数据。

在本实施例中，实际应用数据是独立存在的，例如，应用程序中某个按钮的点击时间、

某个视频的用户点击量等，不受其他类型数据的影响，无法直接的确定哪些数据之间可能存在关联。

因此，在本方案中，通过计算各类型的实际应用数据之间的支持度和置信度，确定置信度较高的数据项中的数据为关联数据，以对关联数据进行关联分析。

在 S104 中，对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

在确定关联数据之后，对关联数据进行关联分析。其中，关联分析的方法可以为线性回归分析、主成分分析、典型相关分析、判别分析、聚类分析、贝叶斯统计分析或者通过统计分析软件进行分析，此处不做限定。

在实际应用中，若将实际应用数据与用户的实际使用习惯或者被测对象的运行方式结合，应用程序中某个按钮的点击时间、某段时间中的点击次数便可以关联起来，以得到应用程序的某个按钮的点击率和点击高峰期等信息，用以评价该按钮对应的功能属性是否符合用户需求或者用户使用率；还可以将网页中某个视频的用户点击量、观看时段或者是重复观看视频段关联起来，以得到这个视频的用户喜好程度、用户类型或者视频关注点。

以上可以看出，本申请实施例提供的一种埋点数据的分析方法通过对被测对象进行埋点获取埋点数据，确定被测对象的实际使用时段，以及该实际使用时段之内的关联数据，通过分析关联数据得到整个实际使用时间之内的用户行为情况和软件的运行状态情况，实现完整、全面的数据统计与分析，提高了用户行为分析的连贯性和可追踪性。

实施例 2

图 2 示出了本申请第二实施例提供的一种埋点数据的分析方法 S102 的具体实现流程图。参见图 2 所示，相对于图 1 所述实施例，本实施例提供的一种埋点数据的分析方法中 S102 包括 S201~S202，具体详述如下：

在 S201 中，根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段。

在实际应用中，由于通过埋点获取到的数据是长时间的数据集，但是大量的数据又会造成较大的数据冗余，浪费数据处理和分析的时间。因此，为了更准确地确定被测对象在被使用过程中的实际使用数据，则根据用户的操作行为确定用户使用被测对象的使用时长。将以下任一时间段确定所述被测对象的实际使用时段：从所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、从所述被测对象启动到被切换到后台之间的时间段，以

及从所述被测对象启动到关闭之间的时间段。

进一步的，由于很多被测对象在退出之后仍旧会在后台运行，或者运行被测对象的终端即使锁屏时候，被测对象也会继续运行，因此，这里设定一个预设时间阈值，用于通过这个时间来截取被测对象的实际使用时段。

需要说明的是，这里的终端可以是手机、平板电脑或者计算机等设备，此处不做限定。

示例性地，请一并结合图 3 所示，图 3 为本实施例提供的实际使用时段的示意图，设定该时间阈值为半分钟。若运行被测对象的终端锁屏超过半分钟，或者半分钟之内没有获取到任何操作信息，则从启动到锁屏的这段时间作为实际使用时段；或者，从被测对象被启动应用到被切换到后台，并且在后台的时间超过半分钟，则视为应用已经使用结束，从启动到切换到后台的半分钟作为实际使用时段；或者，从被测对象被启动到关闭的这段时间作为实际使用时段，但是如果在后台运行超过半分钟；若后台运行时长$t_2 \geq 30$ s，则本次使用时长 t 为：$t = t_1 + t_2 + t_3$，其中，t_1为后台运行之前的前端运行时长，t_3为后台运行之后的前端运行时长；若$t_2 < 30$ s，则本次使用时长 t 为：$t = t_1$，舍弃t_2，t_3则是新一次使用时的使用时长。

需要说明的是，在本实施例中的时间阈值设定为半分钟，还可以根据终端的设置情况、被测对象的运行情况或者用户的使用情况设定该时间阈值，此处不做限定。

在 S202 中，根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

在确定了实际使用时间之后，根据埋点数据的时间标识以及实际使用时间的起点和终点，将时间标识在实际使用时段中的埋点数据识别为实际应用数据。

通过根据用户的操作行为确定用户使用被测对象的使用时长，更准确地确定被测对象在被使用过程中的实际使用数据，降低了数据冗余，提高了数据分析的准确性，进而得到准确的用户行为参数或者应用程序运行情况。

实施例 3

图 4 示出了本申请第三实施例提供的一种埋点数据的分析方法 S103 的具体实现流程图。参见图 4 所示，相对于图 1 所述实施例，本实施例提供的一种埋点数据的分析方法中 S103 包括 S401~S404，具体详述如下：

在 S401 中，根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数。

每个实际应用数据的数据项都有各自的类型标识，该类型可以是按照数据项的属性进

行命名，例如，按钮的点击时间、点击次数或者视频的观看时长等。通过根据实际应用数据的数据项的类型标识，对同一类型标识的数据项进行计数得到数据项数，以通过该数据项数确定实际使用时间中该类数据的发生频率。

在 S402 中，根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度。

本实施例中，第一类型标识对应的数据项为第一数据项，第二类型标识对应的数据项为第二数据项。支持度用于衡量数据项中第一数据项与第二数据项同时出现的频率关系。

具体的，步骤 S402 可以包括：

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A = \frac{support\_count(A)}{support\_count(N)};$$

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A \to B = \frac{support\_count(A \cup B)}{support\_count(N)};$$

其中，$support\_count(A)$ 用于表示第一类型标识对应的第一数据项的数据项数；$support\_count(N)$ 用于表示所有类型标识对应的数据项的总数据项数；$support\_count(A \cup B)$ 用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

示例性地，假如被测的应用程序中的两个按钮分别为第一按钮和第二按钮，在实际应用时间中的按动次数分别记为第一数据项和第二数据项，这两个数据项分别对应第一类型标识与第二类型标识。计算第一数据项的数据项数与总数据项数之比，作为前项支持度。前项支持度用于衡量第一数据项在总数据项中的比重，即该按钮的按动次数在总的按钮按动次数中的比重；确定第一类型标识与第二类型标识共同对应的数据项的数据项数，计算该数据项数与总数据项数之比，作为关联支持度。关联支持度用于衡量第一数据和第二数据项对应的动作共同发生的项数在总按钮按动次数中的比重。

在 S403 中，根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度。

根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A \to B = \frac{Sup\_A \to B}{Sup\_A}。$$

通过用关联支持度与前项支持度相比，确定第一数据项与第二数据项之间的置信度，

置信度用于体现当 A 出现时，B 是否也会出现或有多大概率出现。如果置信度度为 100%，则 A 和 B 是必然的关联关系。如果置信度太低，则说明 A 的出现与 B 是否出现关系不大，即两个数据项之间不存在关联关系。

在 S404 中，将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

在计算出第一数据项和第二数据项之间的置信度之后，通过衡量该置信度的大小确定第一数据项和第二数据项之间的关联关系的强弱。

在本实施例中，设定一个置信度阈值，来衡量第一数据项和第二数据项之间的关联关系的强弱。具体的，若置信度大于或者等于该置信度阈值，则第一数据项和第二数据项中的数据关联程度较强，第一数据项和第二数据项中的数据为所述关联数据；若置信度小于该置信度阈值，则第一数据项和第二数据项中的数据关联程度较若，第一数据项和第二数据项中的数据并不是关联数据。

通过计算不同数据标识对应的数据项之间的置信度，可确定不同数据项之间的关联程度，进而确定关联程度较高的数据项为关联数据项，并对这些数据项之间进行关联分析，保证了埋点数据分析过程和结果的完整与全面。

实施例 4

图 5 示出了本申请实施例四提供的一种终端设备，该装置包括的各单元用于执行图 1 对应的实施例中的各步骤。具体请参阅图 1 与图 1 所对应的实施例中的相关描述。为了便于说明，仅示出了与本实施例相关的部分。

参见图 5，所述终端设备包括：

埋点数据获取单元 501，用于获取被测对象的埋点数据；所述被测对象包括应用程序或网页。

实际数据确定单元 502，用于确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据。

关联数据确定单元 503，用于根据预设模型确定所述实际应用数据中的关联数据。

关联数据分析单元 504，用于对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

可选的，实际数据确定单元 502 可以包括：

使用时段确定单元，用于根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到

关闭之间的时间段确定所述被测对象的实际使用时段。

实际数据识别单元，用于根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

可选的，关联数据确定单元 503 可以包括：

项数技术单元，用于根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数。

支持度单元，用于根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度。

置信度单元，用于根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度。

关联数据单元，用于将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

可选的，支持度单元可以包括：

前项支持度单元，用于根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A = \frac{support\_count(A)}{support\_count(N)};$$

关联支持度单元，用于根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A \to B = \frac{support\_count(A \cup B)}{support\_count(N)};$$

其中，$support\_count(A)$ 用于表示第一类型标识对应的第一数据项的数据项数；$support\_count(N)$ 用于表示所有类型标识对应的数据项的总数据项数；$support\_count(A \cup B)$ 用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

可选的，置信度单元可以包括：

置信度计算单元，用于根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A \to B = \frac{Sup\_A \to B}{Sup\_A}。$$

因此，本申请实施例提供的终端设备可以通过获取埋点数据，确定被测对象的实际使用时段，以及该实际使用时段之内的关联数据，通过分析关联数据得到整个实际使用时间之内的用户行为情况和软件的运行状态情况，实现完整、全面的数据统计与分析，提高了

用户行为分析的连贯性和可追踪性。

实施例 5

图 6 是本申请实施例五提供的一种终端设备的示意图。如图 6 所示，该实施例的终端设备 6 包括：处理器 60、存储器 61 以及存储在所述存储器 61 中并可在所述处理器 60 上运行的计算机程序 62，例如置信度的计算程序。所述处理器 60 执行所述计算机程序 62 时实现上述各个埋点数据的分析方法实施例中的步骤，例如图 1 所示的 S101 至 S104。或者，所述处理器 60 执行所述计算机程序 62 时实现上述各装置实施例中各单元的功能，例如图 5 所示模块 501 至 504 功能。

示例性的，所述计算机程序 62 可以被分割成一个或多个单元，所述一个或者多个单元被存储在所述存储器 61 中，并由所述处理器 60 执行，以完成本申请。所述一个或多个单元可以是能够完成特定功能的一系列计算机程序指令段，该指令段用于描述所述计算机程序 62 在所述终端设备 6 中的执行过程。例如，所述计算机程序 62 可以被分割成埋点数据获取单元、实际数据确定单元、关联数据确定单元以及关联数据分析单元，各单元具体功能如下：

埋点数据获取单元，用于获取被测对象的埋点数据；所述被测对象包括应用程序或网页。

实际数据确定单元，用于确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据。

关联数据确定单元，用于根据预设模型确定所述实际应用数据中的关联数据。

关联数据分析单元，用于对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

所述终端设备可包括，但不仅限于，处理器 60、存储器 61。本领域技术人员可以理解，图 6 仅仅是终端设备 6 的示例，并不构成对终端设备 6 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述终端设备还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 60 可以是中央处理单元（Central Processing Unit，CPU），还可以是其他通用处理器、数字信号处理器 （Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

所述存储器61可以是所述终端设备6的内部存储单元，例如终端设备6的硬盘或内存。所述存储器61也可以是所述终端设备6的外部存储设备，例如所述终端设备6上配备的插接式硬盘，智能存储卡（Smart Media Card，SMC），安全数字（Secure Digital，SD）卡，闪存卡（Flash Card）等。进一步地，所述存储器61还可以既包括所述终端设备6的内部存储单元也包括外部存储设备。所述存储器61用于存储所述计算机程序以及所述终端设备6所需的其他程序和数据。所述存储器61还可以用于暂时地存储已经输出或者将要输出的数据。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内

权利要求书

1. 一种埋点数据的分析方法，其特征在于，包括：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

2. 如权利要求 1 所述的埋点数据的分析方法，其特征在于，所述确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据，包括：

根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段；

根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

3. 如权利要求 1 或 2 所述的埋点数据的分析方法，其特征在于，所述根据预设模型确定所述实际应用数据中的关联数据，包括：

根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数；

根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度；

根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度；

将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

4. 如权利要求 3 所述的埋点数据的分析方法，其特征在于，所述根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度，包括：

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A=\frac{support\_count(A)}{support\_count(N)};$$

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A\rightarrow B=\frac{support\_count(A\cup B)}{support\_count(N)};$$

其中，$support_count(A)$用于表示第一类型标识对应的第一数据项的数据项数；$support_count(N)$用于表示所有类型标识对应的数据项的总数据项数；$support_count(A\cup B)$用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

5. 如权利要求 4 所述的埋点数据的分析方法，其特征在于，所述根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度，包括：

根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A\rightarrow B=\frac{Sup\_A\rightarrow B}{Sup\_A}。$$

6. 一种终端设备，其特征在于，包括：

埋点数据获取单元，用于获取被测对象的埋点数据；所述被测对象包括应用程序或网页。

实际数据确定单元，用于确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据。

关联数据确定单元，用于根据预设模型确定所述实际应用数据中的关联数据。

关联数据分析单元，用于对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

7. 如权利要求 6 所述的终端设备，其特征在于，所述实际数据确定单元包括：

使用时段确定单元，用于根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段。

实际数据识别单元，用于根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

8. 如权利要求 6 或 7 所述的终端设备，其特征在于，所述关联数据确定单元包括：

项数技术单元，用于根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数。

支持度单元，用于根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度。

置信度单元，用于根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度。

关联数据单元，用于将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

9. 如权利要求 8 所述的终端设备，其特征在于，所述支持度单元包括：

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A = \frac{support\_count(A)}{support\_count(N)};$$

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A \to B = \frac{support\_count(A \cup B)}{support\_count(N)};$$

其中，$support\_count(A)$ 用于表示第一类型标识对应的第一数据项的数据项数；$support\_count(N)$ 用于表示所有类型标识对应的数据项的总数据项数；$support\_count(A \cup B)$ 用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

10. 如权利要求 9 所述的埋点数据的分析方法，其特征在于，所述置信度单元包括：

置信度计算单元，用于根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A \to B = \frac{Sup\_A \to B}{Sup\_A}。$$

11. 一种服务器，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

12. 根据权利要求 11 所述的服务器，其特征在于，所述确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据，包括：

根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段；

根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

13. 根据权利要求 11 或 12 所述的服务器，其特征在于，所述根据预设模型确定所述实际应用数据中的关联数据，包括：

根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数；

根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度；

根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度；

将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

14. 根据权利要求 13 所述的服务器，其特征在于，所述根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度，包括：

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A = \frac{support\_count(A)}{support\_count(N)};$$

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A \to B = \frac{support\_count(A \cup B)}{support\_count(N)};$$

其中，$support\_count(A)$ 用于表示第一类型标识对应的第一数据项的数据项数；$support\_count(N)$ 用于表示所有类型标识对应的数据项的总数据项数；$support\_count(A \cup B)$ 用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

15. 根据权利要求 14 所述的服务器，其特征在于，所述根据所述第一数据项与所述

第二数据项之间的支持度计算其对应的置信度，包括：

根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A \rightarrow B = \frac{Sup\_A \rightarrow B}{Sup\_A}$$。

16. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：

获取被测对象的埋点数据；所述被测对象包括应用程序或网页；

确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据；

根据预设模型确定所述实际应用数据中的关联数据；

对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息。

17. 根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据，包括：

根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段；

根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据。

18. 根据权利要求 16 或 17 所述的计算机可读存储介质，其特征在于，所述根据预设模型确定所述实际应用数据中的关联数据，包括：

根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数；

根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度；

根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度；

将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据。

19. 根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二

数据项之间的支持度，包括：

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的前项支持度：

$$Sup\_A = \frac{support\_count(A)}{support\_count(N)};$$

根据如下公式计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的关联支持度：

$$Sup\_A \to B = \frac{support\_count(A \cup B)}{support\_count(N)};$$

其中，$support\_count(A)$ 用于表示第一类型标识对应的第一数据项的数据项数；$support\_count(N)$ 用于表示所有类型标识对应的数据项的总数据项数；$support\_count(A \cup B)$ 用于表示第一类型标识与第二类型标识共同对应的数据项的数据项数。

20. 根据权利要求 19 所述的计算机可读存储介质，其特征在于，所述根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度，包括：

根据如下公式计算第一数据项与所述第二数据项之间的置信度：

$$Conf\_A \to B = \frac{Sup\_A \to B}{Sup\_A}。$$

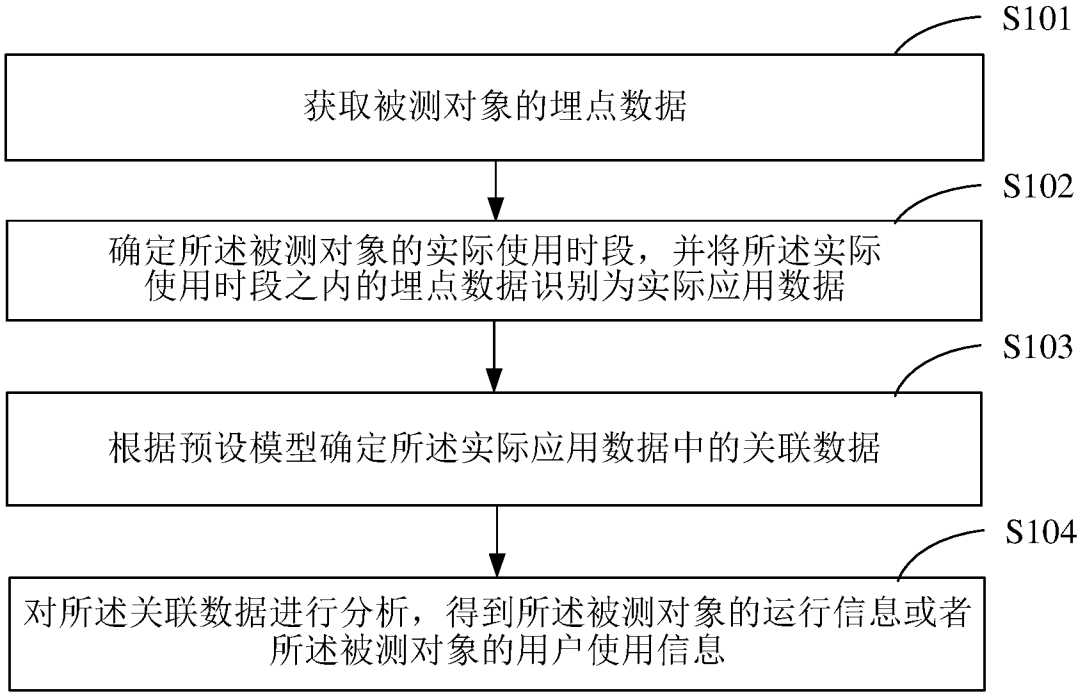
S101
获取被测对象的埋点数据
S102
确定所述被测对象的实际使用时段，并将所述实际使用时段之内的埋点数据识别为实际应用数据
S103
根据预设模型确定所述实际应用数据中的关联数据
S104
对所述关联数据进行分析，得到所述被测对象的运行信息或者所述被测对象的用户使用信息

图 1

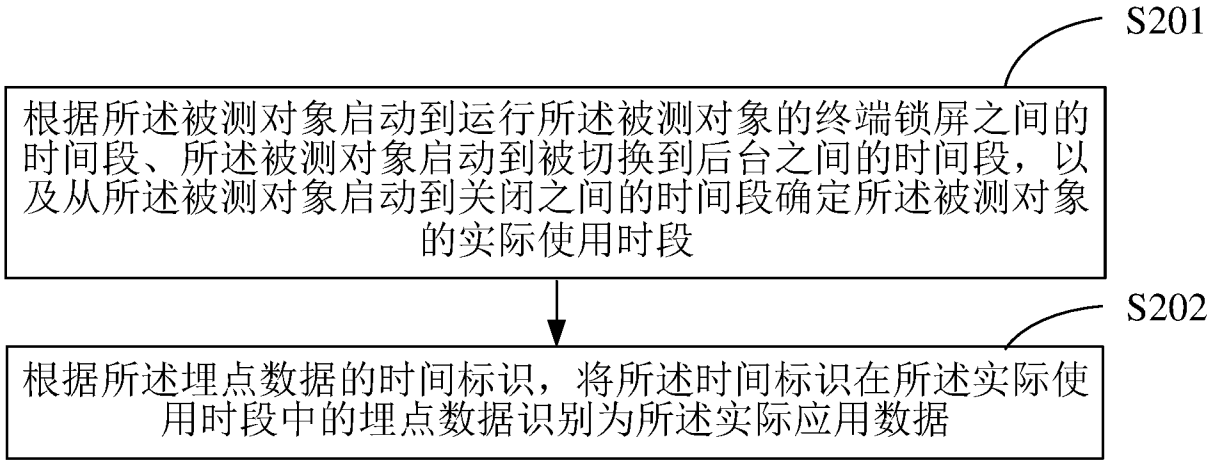
S201
根据所述被测对象启动到运行所述被测对象的终端锁屏之间的时间段、所述被测对象启动到被切换到后台之间的时间段，以及从所述被测对象启动到关闭之间的时间段确定所述被测对象的实际使用时段
S202
根据所述埋点数据的时间标识，将所述时间标识在所述实际使用时段中的埋点数据识别为所述实际应用数据

图 2

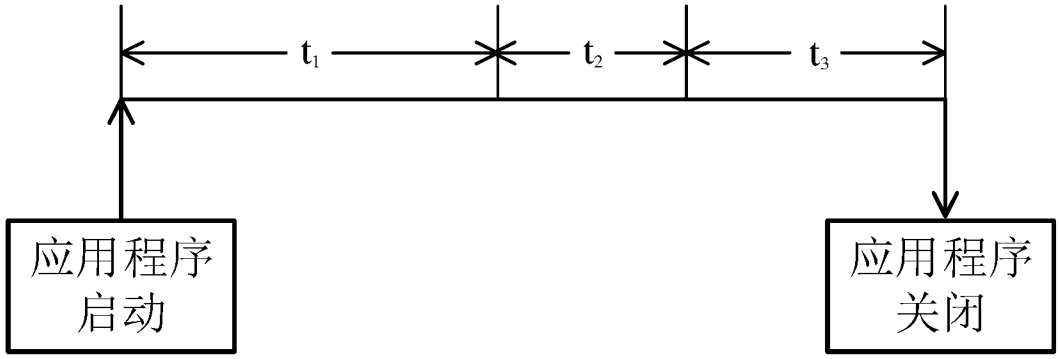
t1
t2
t3
应用程序启动
应用程序关闭

图 3

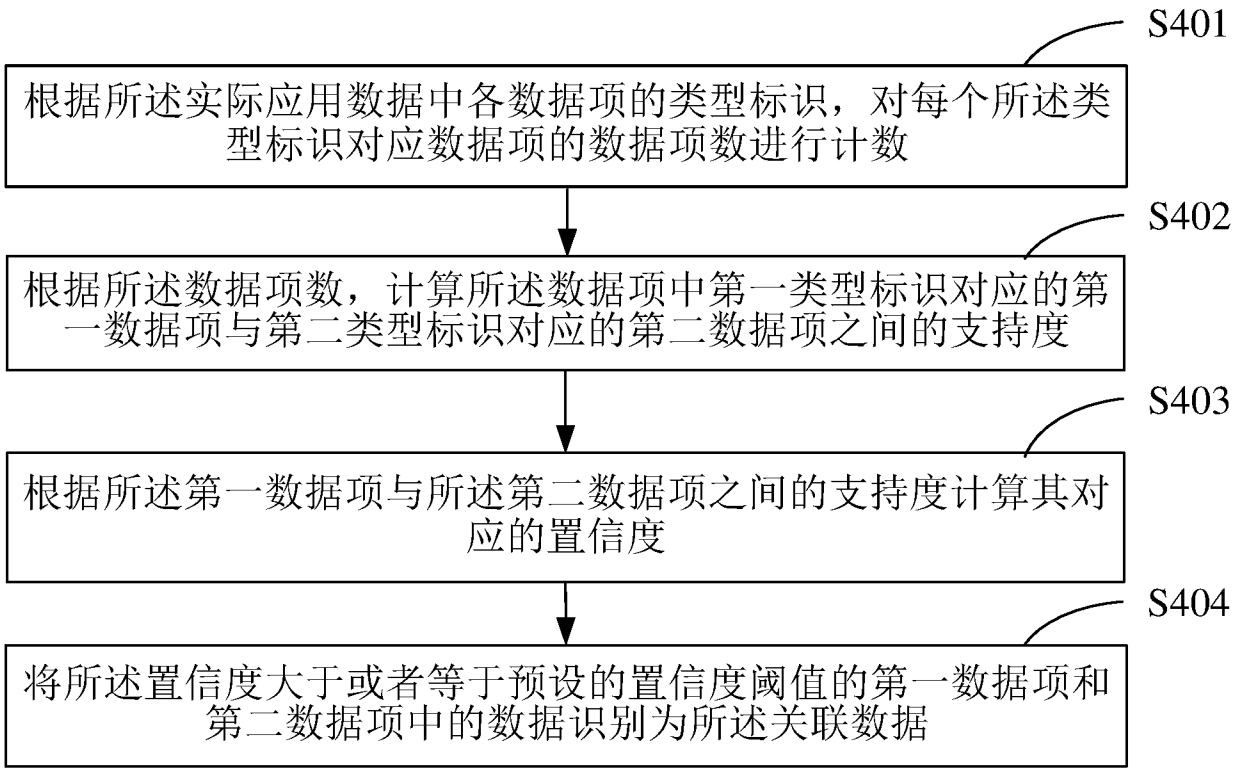
S401
根据所述实际应用数据中各数据项的类型标识，对每个所述类型标识对应数据项的数据项数进行计数
S402
根据所述数据项数，计算所述数据项中第一类型标识对应的第一数据项与第二类型标识对应的第二数据项之间的支持度
S403
根据所述第一数据项与所述第二数据项之间的支持度计算其对应的置信度
S404
将所述置信度大于或者等于预设的置信度阈值的第一数据项和第二数据项中的数据识别为所述关联数据

图 4

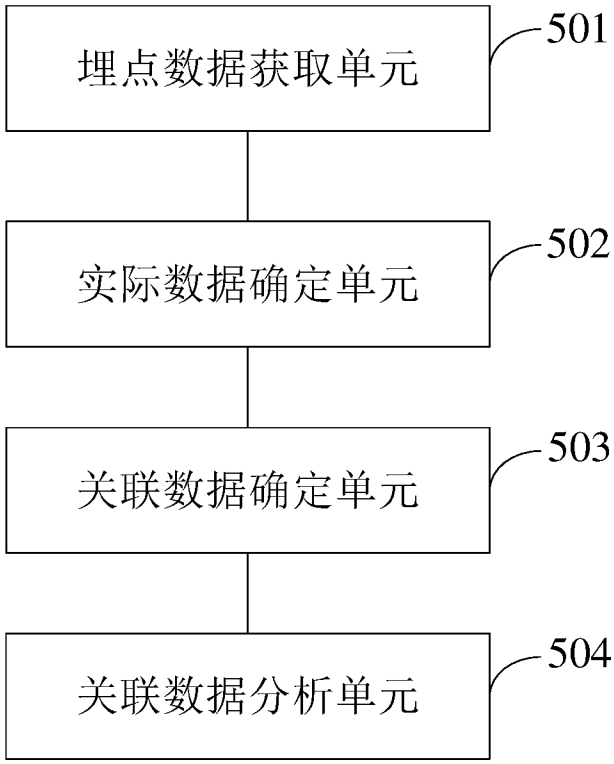
埋点数据获取单元
501
实际数据确定单元
502
关联数据确定单元
503
关联数据分析单元
504

图 5

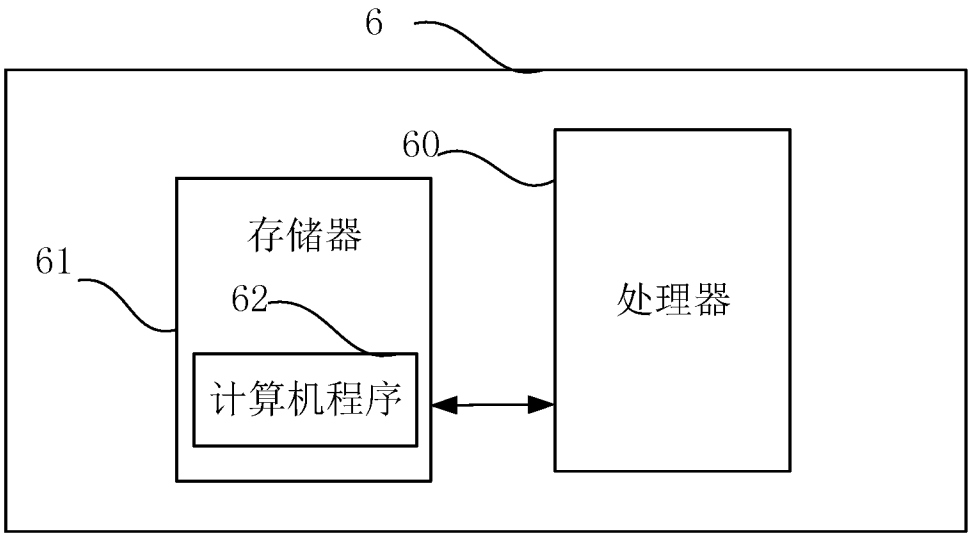
6
60
存储器
61
62
计算机程序
处理器

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/097448

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/34(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: 埋点, 用户, 行为, 点击, 时间, 期间, 相关, 关联, 支持度, 置信度, track+, user, click+, behavior, operat+, time, period, relat+, associat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102946319 A (FOCUS TECHNOLOGY CO., LTD. ET AL.) 27 February 2013 (2013-02-27) description, paragraphs [0031]-[0036]	1-2, 6-7, 11-12, 16-17
A	CN 104572043 A (ALIBABA GROUP HOLDING LIMITED) 29 April 2015 (2015-04-29) entire document	1-20
A	CN 107818162 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 20 March 2018 (2018-03-20) entire document	1-20
A	CN 107220745 A (BEIJING HONGMA MEDIA CULTURE DEVELOPMENT CO., LTD.) 29 September 2017 (2017-09-29) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 February 2019	**27 February 2019**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing** **100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/097448

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	102946319	A	27 February 2013	None	
CN	104572043	A	29 April 2015	None	
CN	107818162	A	20 March 2018	None	
CN	107220745	A	29 September 2017	None	

国际检索报告

国际申请号
PCT/CN2018/097448

A. 主题的分类

G06F 11/34(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNKI, CNPAT, WPI, EPODOC: 埋点，用户，行为，点击，时间，期间，相关，关联，支持度，置信度，track+, user, click+, behavior, operat+, time, period, relat+, associat+

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 102946319 A (焦点科技股份有限公司 等) 2013年 2月 27日 (2013 - 02 - 27) 说明书第[0031]-[0036]段	1-2，6-7， 11-12，16-17
A	CN 104572043 A (阿里巴巴集团控股有限公司) 2015年 4月 29日 (2015 - 04 - 29) 全文	1-20
A	CN 107818162 A (平安科技深圳有限公司) 2018年 3月 20日 (2018 - 03 - 20) 全文	1-20
A	CN 107220745 A (北京红马传媒文化发展有限公司) 2017年 9月 29日 (2017 - 09 - 29) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 2月 18日	2019年 2月 27日

ISA/CN的名称和邮寄地址	受权官员
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	王伟
传真号 (86-10)62019451	电话号码 86-(10)-53961525

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/097448

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	102946319	A	2013年 2月 27日	无	
CN	104572043	A	2015年 4月 29日	无	
CN	107818162	A	2018年 3月 20日	无	
CN	107220745	A	2017年 9月 29日	无	