



(12) 发明专利申请

(10) 申请公布号 CN 113934616 A

(43) 申请公布日 2022. 01. 14

(21) 申请号 202111540706.3

(22) 申请日 2021.12.16

(71) 申请人 深圳市活力天汇科技股份有限公司

地址 518057 广东省深圳市南山区粤海街道高新南九道10号深圳湾科技生态园10栋B座13层01-08号

(72) 发明人 刘小雷 李尚锦

(51) Int. Cl.

G06F 11/34 (2006.01)

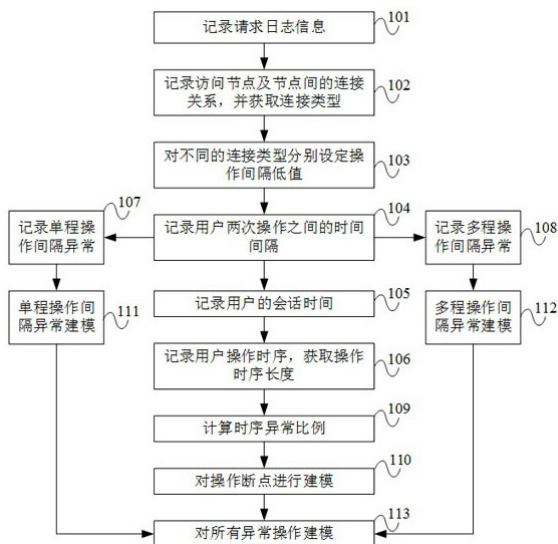
权利要求书3页 说明书6页 附图1页

(54) 发明名称

一种基于用户操作时序判断异常用户的方法

(57) 摘要

本发明提供了一种基于用户操作时序判断异常用户的方法,通过记录本次请求的用户、请求时间、访问页面及访问节点。判断连接类型,获取操作间隔值及操作时序。判断是否有操作断点,若有则记录所述操作断点的个数。判断操作间隔值是否低于预设的操作间隔低值。对操作断点进行操作断点异常建模,根据连接类型建立单程操作间隔异常建模和多程操作间隔异常建模。根据不同操作类型的操作时序、各时序对应的权值,根据计算最终异常指标F。优点是,在机票搜索业务中,实时记录用户操作过程,通过跟踪操作异常节点,对操作过程进行加权处理,从而识别异常用户,以供后续处理。



1. 一种基于用户操作时序判断异常用户的方法,其特征在于,所述方法包括:
 记录本次请求的用户、请求时间、访问页面及访问节点;
 根据访问页面及访问节点构成的访问路径判断连接类型,获取操作间隔值及操作时序;
 根据所述操作间隔值采集操作断点;
 对所述操作断点进行操作断点异常建模,具体为,对多个时间段内的每个操作时序的操作断点比例进行加权;
 根据连接类型建立单程操作间隔异常建模和多程操作间隔异常建模;
 根据不同操作类型的操作时序、各时序对应的权值,根据下式计算出最终异常指标F:

$$F = j_1 F_1 + j_2 F_2 + j_3 F_3 ;$$

若F超过阈值,可以将此用户判断为异常用户,其中j为加权参数,表示此类异常的重要程度, F_1 为在不考虑连接类型的情况下的操作断点加权后权值, F_2 为单程操作中每个操作时序内的单程操作间隔的异常比例加权相加值, F_3 为多程操作中每个操作时序内的多程操作间隔的异常比例加权相加值; j_1 为 F_1 的重要性加权参数; j_2 为 F_2 的重要性加权参数; j_3 为 F_3 的重要性加权参数。

2. 根据权利要求1所述的基于用户操作时序判断异常用户的方法,其特征在于,所述判断连接类型,获取操作间隔值及操作时序,包括:

若连接中的相邻被操作页面间页面序号连续,则为单程连接;
 若连接中的相邻被操作页面间页面序号不连续,则为多程连接;
 将用户在一个周期内的所有操作,划分为若干个操作时序。

3. 根据权利要求1所述的基于用户操作时序判断异常用户的方法,其特征在于,所述根据所述操作间隔值采集操作断点,包括:

判断相邻操作的所述操作间隔值是否低于预设的操作间隔低值,若是则异常,否则正常;

若异常则认为当前操作时序中此操作为操作断点,记录所述操作断点。

4. 根据权利要求1所述的基于用户操作时序判断异常用户的方法,其特征在于,所述对多个时间段内的每个操作时序的操作断点比例进行加权,包括:

$$F_1 = h_1 \frac{x_1}{N_1} + h_2 \frac{x_2}{N_2} + h_3 \frac{x_3}{N_3} + \dots + h_n \frac{x_n}{N_n} ;$$

其中,权值 F_1 为在不考虑连接类型的情况下的操作断点加权后权值,加权参数h为根据操作时序长度N乘以[0,1]区间值得到的,用以表示此操作时序的重要程度; x_1 为在不考虑连接类型情况下第1个时序中的操作断点个数; x_2 为在不考虑连接类型情况下第2个时序中的操作断点个数; x_3 为在不考虑连接类型情况下第3个时序中的操作断点个数; x_n 为在不考虑连接类型情况下第n个时序中的操作断点个数; N_1 为第1个时序的长度; N_2

为第2个时序的长度； N_3 为第3个时序的长度； N_n 为第n个时序的长度； h_1 为第1个时序的加权参数； h_2 为第2个时序的加权参数； h_3 为第3个时序的加权参数； h_n 为第n个时序的加权参数。

5. 根据权利要求2所述的基于用户操作时序判断异常用户的方法，其特征在于，所述根据所述操作间隔值采集操作断点，包括：

若用户的操作间隔低于预设操作间隔值时，则此次操作为异常，对异常操作计数；根据异常操作的总量和所述时序长度计算异常操作比例。

6. 根据权利要求5所述的基于用户操作时序判断异常用户的方法，其特征在于，所述对所述操作断点进行断点异常建模，根据一周期内每个操作时序的操作断点比例进行加权，包括：

$$\text{当前连接为单程连接时，} F_2 = h_1 \frac{x_1'}{N_1} + h_2 \frac{x_2'}{N_2} + h_3 \frac{x_3'}{N_3} + \dots + h_n \frac{x_n'}{N_n} ;$$

加权参数h为根据操作时序长度N乘以[0, 1]区间值得到的，x为异常操作次数； x_1' 为当前连接模式为单程连接的情况下第1个时序中的操作断点个数， x_2' 为当前连接模式为单程连接的情况下第2个时序中的操作断点个数， x_3' 为当前连接模式为单程连接的情况下第3个时序中的操作断点个数， x_n' 为当前连接模式为单程连接的情况下第n个时序中的操作断点个数； N_1 为第一个时序的长度， N_2 为第2个时序的长度， N_3 为第3个时序的长度， N_n 为第n个时序的长度； h_1 为第1个时序的加权参数； h_2 为第2个时序的加权参数； h_3 为第3个时序的加权参数； h_n 为第n个时序的加权参数。

7. 根据权利要求5所述的基于用户操作时序判断异常用户的方法，其特征在于，所述对所述操作断点进行断点异常建模，根据一周期内每个操作时序的操作断点比例进行加权，包括：

$$\text{当前连接为多程连接时，} F_3 = h_1 \frac{x_1''}{N_1} + h_2 \frac{x_2''}{N_2} + h_3 \frac{x_3''}{N_3} + \dots + h_n \frac{x_n''}{N_n} ;$$

加权参数h为根据操作时序长度N乘以[0, 1]区间值得到的，x为异常操作次数； x_1'' 为当前连接模式为多程连接的情况下第1个时序中的操作断点个数； x_2'' 为当前连接模式为多程连接的情况下第2个时序中的操作断点个数； x_3'' 为当前连接模式为多程连接的情况下第3个时序中的操作断点个数； x_n'' 为当前连接模式为多程连接的情况下第n个时序的操作断点个数； N_1 为第1个时序的长度； N_2 为第2个时序的长度； N_3 为第3个时序的长度； N_n 为第n个时序的长度； h_1 为第1个时序的加权参数； h_2 为第2个时序的加权参数； h_3

为第3个时序的加权参数; h_n 为第n个时序的加权参数。

一种基于用户操作时序判断异常用户的方法

技术领域

[0001] 本发明涉及计算机软件领域,尤其涉及一种基于用户操作时序判断异常用户的方法。

背景技术

[0002] 在机票搜索领域中,服务器除了接收正常用户的请求之外,往往会收到机器人等异常用户的大量请求。这些请求并不会最终产生收益,最好将其识别以做后续处理。所以如何提供一种能够自动识别异常用户请求的方法成为亟待解决的问题。

发明内容

[0003] 本发明提供一种基于用户操作时序判断异常用户的方法,用以解决现有技术中无法在机票搜索领域中通过分析用户的行为,判断此用户是否为异常用户的问题。

[0004] 为了实现上述目的,本发明技术方案提供了一种基于用户操作时序判断异常用户的方法,包括:记录本次请求的用户、请求时间、访问页面及访问节点。根据访问页面及访问节点构成的访问路径判断连接类型,获取操作间隔值及操作时序。判断当前操作时序中是否有操作断点,对操作断点进行操作断点异常建模,具体为,对多个时间段内的每个操作时序的操作断点比例进行加权。根据连接类型建立单程操作间隔异常建模和多程操作间隔异常建模。根据不同操作类型的操作时序、各时序对应的权值,根据下式计算出最终异常指标

$F: F = j_1 F_1 + j_2 F_2 + j_3 F_3$; 若F超过阈值,可以将此用户判断为异常用户,其中j为加权参数,表示此类异常的重要程度, F_1 为在不考虑连接类型的情况下的操作断点加权后权值, F_2 为单程操作中每个操作时序内的单程操作间隔的异常比例加权相加值, F_3 为多程操作中每个操作时序内的多程操作间隔的异常比例加权相加值; j_1 为 F_1 的重要性加权参数; j_2 为 F_2 的重要性加权参数; j_3 为 F_3 的重要性加权参数。

[0005] 作为上述技术方案的优选,较佳的,判断连接类型,获取操作间隔值及操作时序,包括:若连接中的相邻被操作页面间页面序号连续,则为单程连接;若连接中的相邻被操作页面间页面序号不连续,则为多程连接;将用户在一个周期内的所有操作,划分为若干个操作时序。

[0006] 作为上述技术方案的优选,较佳的,根据操作间隔值采集操作断点,包括:判断相邻操作的所述操作间隔值是否低于预设的操作间隔低值,若是则异常,否则正常;若异常则认为当前操作时序中此操作为操作断点,记录操作断点。

[0007] 作为上述技术方案的优选,较佳的,对多个时间段内的每个操作时序的操作断点比例进行加权,包括:

$$F_1 = h_1 \frac{x_1}{N_1} + h_2 \frac{x_2}{N_2} + h_3 \frac{x_3}{N_3} + \dots + h_n \frac{x_n}{N_n};$$

其中,权值 F_1 为在不考虑连接类型的情况下的操作断点加权后权值,加权参数 h 为根据操作时序长度 N 乘以 $[0, 1]$ 区间值得到的,用以表示此操作时序的重要程度; x_1 为在不考虑连接类型情况下第1个时序中的操作断点个数; x_2 为在不考虑连接类型情况下第2个时序中的操作断点个数; x_3 为在不考虑连接类型情况下第3个时序中的操作断点个数; x_n 为在不考虑连接类型情况下第 n 个时序中的操作断点个数; N_1 为第1个时序的长度; N_2 为第2个时序的长度; N_3 为第3个时序的长度; N_n 为第 n 个时序的长度; h_1 为第1个时序的加权参数; h_2 为第2个时序的加权参数; h_3 为第3个时序的加权参数; h_n 为第 n 个时序的加权参数。

[0008] 作为上述技术方案的优选,较佳的,根据所述操作间隔值采集操作断点,包括:若用户的操作间隔低于预设操作间隔值时,则此次操作为异常,对异常操作计数;

根据异常操作的总量和所述时序长度计算异常操作比例。

[0009] 作为上述技术方案的优选,较佳的,对所述操作断点进行操作断点异常建模,根据一周期内每个操作时序的操作断点比例进行加权,包括:

$$\text{当前连接为单程连接时, } F_2 = h_1 \frac{x_1'}{N_1} + h_2 \frac{x_2'}{N_2} + h_3 \frac{x_3'}{N_3} + \dots + h_n \frac{x_n'}{N_n};$$

加权参数 h 为根据操作时序长度 N 乘以 $[0, 1]$ 区间值得到的, x 为异常操作次数; x_1' 为当前连接模式为单程连接的情况下第1个时序中的操作断点个数, x_2' 为当前连接模式为单程连接的情况下第2个时序中的操作断点个数, x_3' 为当前连接模式为单程连接的情况下第3个时序中的操作断点个数, x_n' 为当前连接模式为单程连接的情况下第 n 个时序中的操作断点个数; N_1 为第一个时序的长度, N_2 为第2个时序的长度, N_3 为第3个时序的长度, N_n 为第 n 个时序的长度; h_1 为第1个时序的加权参数; h_2 为第2个时序的加权参数; h_3 为第3个时序的加权参数; h_n 为第 n 个时序的加权参数。

[0010] 作为上述技术方案的优选,较佳的,对所述操作断点进行操作断点异常建模,根据一周期内每个操作时序的操作断点比例进行加权,包括:

$$\text{当前连接为多程连接时, } F_3 = h_1 \frac{x_1''}{N_1} + h_2 \frac{x_2''}{N_2} + h_3 \frac{x_3''}{N_3} + \dots + h_n \frac{x_n''}{N_n};$$

加权参数 h 为根据操作时序长度 N 乘以 $[0, 1]$ 区间值得到的, x 为异常操作次数; x_1'' 为当前连接模式为多程连接的情况下第1个时序中的操作断点个数; x_2'' 为当前连

接模式为多程连接的情况下第2个时序中的操作断点个数； x_3'' 为当前连接模式为多程连接的情况下第3个时序中的操作断点个数； x_n'' 为当前连接模式为多程连接的情况下第n个时序的操作断点个数； N_1 为第1个时序的长度； N_2 为第2个时序的长度； N_3 为第3个时序的长度； N_n 为第n个时序的长度； h_1 为第1个时序的加权参数； h_2 为第2个时序的加权参数； h_3 为第3个时序的加权参数； h_n 为第n个时序的加权参数。

[0011] 本发明技术方案提供的，所述方法包括：本发明提供了一种基于用户操作时序判断异常用户的方法，通过记录本次请求的用户、请求时间、访问页面及访问节点。根据访问页面及访问节点构成的访问路径判断连接类型，采集操作间隔值及操作时序。判断当前操作时序中是否有操作断点，若有则记录所述操作断点的个数。判断操作间隔值是否低于预设的操作间隔低值，若是则异常，否则正常。对操作断点进行异常建模，具体为，对多个时间段内的每个操作时序的操作断点比例进行加权。根据连接类型建立单程操作间隔异常建模和多程操作间隔异常建模。根据不同操作类型的操作时序、各时序对应的权值，根据下式计算出最终异常指标F： $F = j_1F_1 + j_2F_2 + j_3F_3$ ；若F超过阈值，可以将此用户判断为异常用户，其中j为预设值。

[0012] 本发明的优点是：在机票搜索业务中，实时记录用户操作过程，通过跟踪操作异常节点，对操作过程进行加权处理，从而识别异常用户，以供后续处理。

附图说明

[0013] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0014] 图1为本发明技术方案提供的流程图。

[0015] 图2为本发明实施例提供的页面跳转示意图。

具体实施方式

[0016] 为使本发明实施例的目的、技术方案和优点更加清楚，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、完整的描述，显然，所描述的实施例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，本领域普通技术人员在没有做出创造性劳动的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。

[0017] 现结合具体实施方式对本发明进行说明，图1为本发明实施例提供的流程示意图，如图1所示，包括：

步骤101、记录请求日志信息。

[0018] 服务器接收到终端发出的机票搜索请求时，会记录这次请求的用户信息、请求时间、访问页面等动作。

[0019] 步骤102、记录访问节点及节点间的连接关系,并获取连接类型。

[0020] 具体的,用户在某一时间访问的各个页面(首页、列表页、详情页)之间是具有特定的双向连接,双向连接用以表示用户进入当前页面或返回上一页面或进行与页面跳转动作。特殊的,部分页面之间并不存在连接,现详细说明,如图2所示。图2所示中包括单程连接,如P1(页面1)与P2有连接,即表明P1可以进入P2,P2可以返回到P1,但P1无法进入P2之外的页面;图2所示中包括多程连接,P3与P2、P6之间均有连接关系,在此连接关系下,P3返回P2再进入P6,由于页面缓存等因素,可以视为P3与P6存在连接。

[0021] 步骤103、对不同的连接类型分别设定操作间隔低值。

[0022] 根据多数用户操作统计可预估一个最小的操作时间间隔,针对两类连接分别设定两个最小值 m_1 和 m_2 。其中, m_1 为单程操作的预设的操作间隔低值, m_2 为多程操作的预设的操作间隔低值。

[0023] 步骤104、记录用户两次操作之间的时间间隔。

[0024] 判断相邻操作之间的操作间隔值是否低于预设的操作间隔低值,若是,则操作异常,执行步骤107或步骤108,否则正常。其中,操作间隔值是根据两次的操作时间计算得到的。

[0025] 若连接类型为单程连接则在步骤107之后执行步骤111,若连接类型为多程连接则在步骤108之后执行步骤112。

[0026] 步骤105、记录用户的会话时间。

[0027] 会话时间指的是用户进行多次操作的累积时长,在此累积时长内的相邻操作的时间间隔均小于操作间隔低值。若此次操作距离上一次操作的操作间隔大于操作间隔低值,则为此次操作作为新的会话时间起点。

[0028] 步骤106、记录用户操作时序,获取操作时序长度。

[0029] 具体示例如图2所示,记录用户在一个会话时间内进行页面跳转的顺序,例如多程连接(P1,P2,P3,P6),表示用户首先进入P1,进入P2再进入P3,返回至P2后进入P6。

[0030] 若为单程连接则为(P1,P2,P3,P4,P3),表示用户首先进入P1,进入P2再进入P3,再进入P4后返回P3。

[0031] 一个操作时序内,N个节点表明了N次操作,即长度为N。

[0032] 步骤104后开始记录操作断点。

[0033] 结合图2所示,若用户的操作时序为(P1,P2,P6,P4,P8),因P6和P4之间并无连接,无法直接跳转,视(P6,P4)为一个操作断点,此次操作异常。

[0034] 步骤107、记录单程操作间隔异常。

[0035] 若一个用户的某次单程操作间隔低于预设的操作间隔低值 m_1 ,视此次操作为异常。步骤107后执行步骤111。

[0036] 步骤108、记录多程操作间隔异常。

[0037] 若一个用户的某次多程操作间隔低于预设的操作间隔低值 m_2 ,视此次操作为异常。步骤108后执行步骤112。

[0038] 步骤109、计算时序异常比例。

[0039] 一个操作时序里,x个操作异常表明了不符合常理操作有x次,因此异常操作比例为: x/N ,N为操作时序长度。

[0040] 步骤110、对操作断点进行建模。

[0041] 用户可能在一天内多个时间段进行操作,即在一个时间范围内有多个操作时序,以每个操作时序内的操作断点比例加权相加可得:

$$F_1 = h_1 \frac{x_1}{N_1} + h_2 \frac{x_2}{N_2} + h_3 \frac{x_3}{N_3} + \dots + h_n \frac{x_n}{N_n}$$

其中, F_1 为在不考虑连接类型的情况下的操作断点加权后权值,加权参数 h 为根据操作时序长度 N 乘以 $[0, 1]$ 区间值得到的,用以表示此操作时序的重要程度;步骤110的公式中: x_1 为在不考虑连接类型情况下第1个时序中的操作断点个数; x_2 为在不考虑连接类型情况下第2个时序中的操作断点个数; x_3 为在不考虑连接类型情况下第3个时序中的操作断点个数; x_n 为在不考虑连接类型情况下第 n 个时序中的操作断点个数; N_1 为第1个时序的长度; N_2 为第2个时序的长度; N_3 为第3个时序的长度; N_n 为第 n 个时序的长度; h_1 为第1个时序的加权参数; h_2 为第2个时序的加权参数; h_3 为第3个时序的加权参数; h_n 为第 n 个时序的加权参数。

[0042] 步骤111、单程操作间隔异常建模。

[0043] 在单程操作中,以每个操作时序内的单程操作间隔的异常比例加权相加:

$$F_2 = h_1 \frac{x_1'}{N_1} + h_2 \frac{x_2'}{N_2} + h_3 \frac{x_3'}{N_3} + \dots + h_n \frac{x_n'}{N_n}$$

F_2 为单程操作中每个操作时序内的单程操作间隔的异常比例加权相加值,在步骤111的公式中: x_1' 为当前连接模式为单程连接的情况下第1个时序中的操作断点个数, x_2' 为当前连接模式为单程连接的情况下第2个时序中的操作断点个数, x_3' 为当前连接模式为单程连接的情况下第3个时序中的操作断点个数, x_n' 为当前连接模式为单程连接的情况下第 n 个时序中的操作断点个数; N_1 为第一个时序的长度, N_2 为第2个时序的长度, N_3 为第3个时序的长度, N_n 为第 n 个时序的长度; h_1 为第1个时序的加权参数;

h_2 为第2个时序的加权参数; h_3 为第3个时序的加权参数; h_n 为第 n 个时序的加权参数。

[0044] 步骤112、多程操作间隔异常建模。

[0045] 在多程操作中,每个操作时序内的多程操作间隔的异常比例加权相加可得:

$$F_3 = h_1 \frac{x_1''}{N_1} + h_2 \frac{x_2''}{N_2} + h_3 \frac{x_3''}{N_3} + \dots + h_n \frac{x_n''}{N_n}$$

F_3 为多程操作中每个操作时序内的多程操作间隔的异常比例加权相加值； x_1'' 为当前连接模式为多程连接的情况下第1个时序中的操作断点个数； x_2'' 为当前连接模式为多程连接的情况下第2个时序中的操作断点个数； x_3'' 为当前连接模式为多程连接的情况下第3个时序中的操作断点个数； x_n'' 为当前连接模式为多程连接的情况下第n个时序的操作断点个数； N_1 为第1个时序的长度； N_2 为第2个时序的长度； N_3 为第3个时序的长度； N_n 为第n个时序的长度； h_1 为第1个时序的加权参数； h_2 为第2个时序的加权参数； h_3 为第3个时序的加权参数； h_n 为第n个时序的加权参数。

[0046] 步骤113,对所有异常操作建模。

[0047] 在多程操作中,根据异常操作的重要程度分别对三类异常建模进行加权相加可得:

$$F = j_1 F_1 + j_2 F_2 + j_3 F_3$$

其中加权参数j为根据异常的不同重要性设置的,用以表示此类异常的重要程度。最后计算的F为异常指标,表示此时间范围内的用户操作异常性,当F超过阈值则表示此用户为非正常用户的概率越高。

[0048] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,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,或者对其中部分或者全部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发明各实施例技术方案的范围。

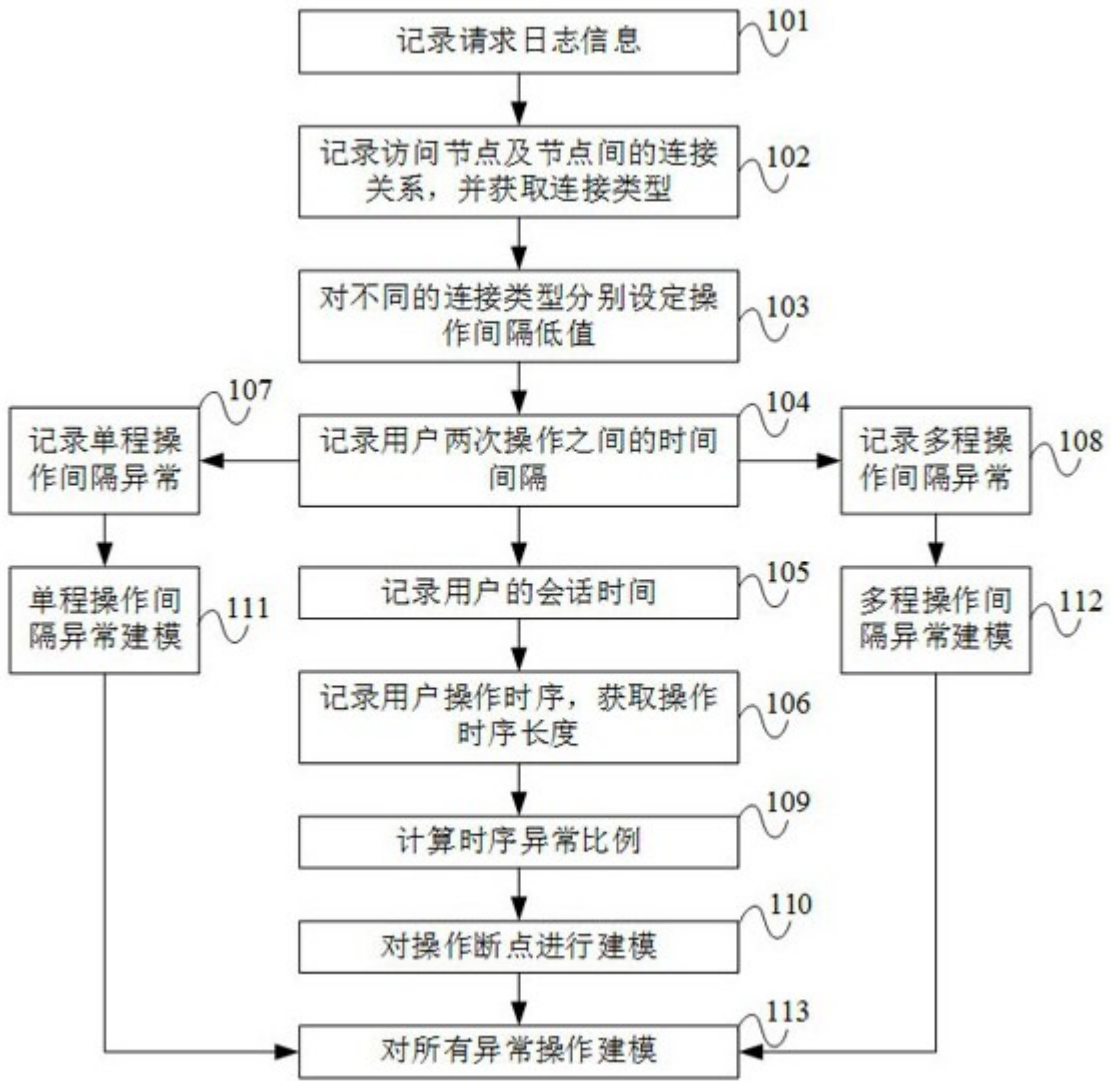


图1

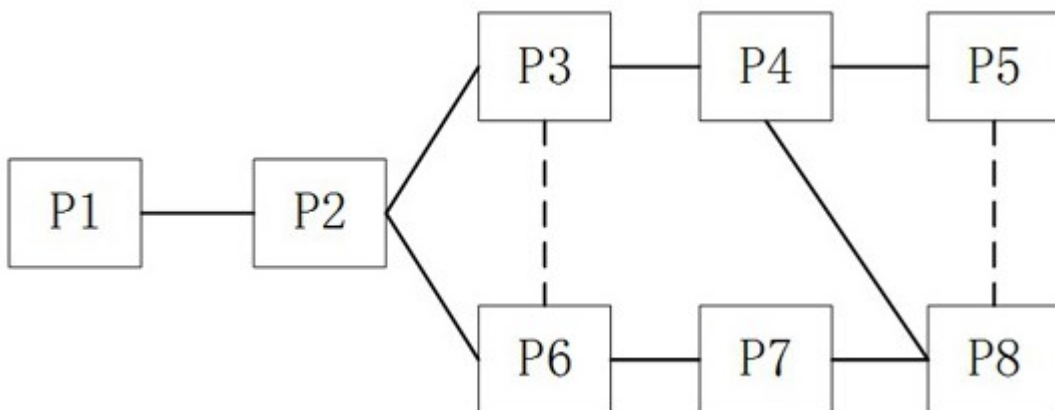


图2