

[19] 中华人民共和国国家知识产权局



# [12] 发明专利申请公布说明书

[21] 申请号 200610061206.0

[51] Int. Cl.

H04L 12/58 (2006.01)

G06Q 10/00 (2006.01)

G06F 17/30 (2006.01)

H04L 9/32 (2006.01)

H04M 3/42 (2006.01)

[43] 公开日 2007 年 11 月 28 日

[11] 公开号 CN 101079824A

[22] 申请日 2006.6.15

[21] 申请号 200610061206.0

[71] 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市高新科技园南区  
高新南一道飞亚达高科技大厦 5-10 层

[72] 发明人 丁江伟

[74] 专利代理机构 深圳市顺天达专利商标代理有限公司

代理人 郭伟刚 蔡晓红

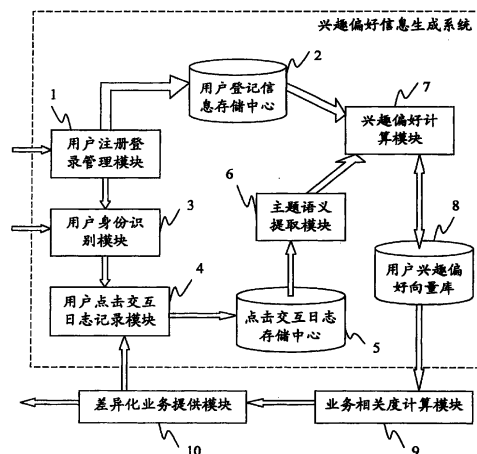
权利要求书 2 页 说明书 7 页 附图 2 页

## [54] 发明名称

一种用户兴趣偏好向量生成系统和方法

## [57] 摘要

本发明提供了一种用户兴趣偏好向量生成系统,包括用户登记信息存储中心、点击交互日志存储中心、以及用于对点击交互日志存储中心中存储的一段时间内的用户点击交互日志进行分析,得到一个按语义强度排序的主题分类标识向量的主题语义提取模块;用于从用户登记信息存储中心获取用户静态登记信息、从主题语义提取模块获取所述主题分类标识向量,并对上述信息进行分析,得到相关用户的兴趣偏好向量的兴趣偏好计算模块。本发明还提供了一种用户兴趣偏好向量生成方法。通过本发明可得到用户的兴趣偏好向量。服务平台可以根据用户的兴趣偏好向量有针对性、有选择地推送信息或者提供服务,突出用户感兴趣的信息和服务。



1、一种用户兴趣偏好向量生成系统，包括，用户登记信息存储中心，其用于存储用户的静态属性；点击交互日志存储中心，其用于记录用户一段时间内的点击交互日志；其特征在于，还包括：主题语义提取模块和兴趣偏好计算模块；

所述主题语义提取模块用于对所述点击交互日志存储中心中存储的一段时间内的用户点击交互日志进行分析，得到一个按语义强度排序的主题分类标识向量；

所述兴趣偏好计算模块用于从所述用户登记信息存储中心获取用户静态登记信息、从所述主题语义提取模块获取所述主题分类标识向量，并对上述信息进行分析，得到相关用户的兴趣偏好向量。

2、根据权利要求1所述的一种用户兴趣偏好向量生成系统，其特征在于：还包括用户兴趣偏好向量库，其用于存储用户的兴趣偏好向量；相应的，所述兴趣偏好计算模块还同时从所述用户兴趣偏好向量库中获取用户的历史兴趣偏好向量，与所述静态登记信息和主题分类标识向量一起进行分析，以得到用户当前的兴趣偏好向量，并存储到所述用户兴趣偏好向量库中。

3、根据权利要求1所述的一种用户兴趣偏好向量生成系统，其特征在于，还包括：

用户注册登录管理模块，其用于处理用户的注册注销、注册信息的修改、用户的身份认证，管理所述用户登记信息存储中心，并为用户身份识别模块提供用户数据；

用户身份识别模块，其用于对用户进行身份识别；

用户点击交互日志记录模块，其用于记录用户本次在服务平台上的行为，并存储到所述点击交互日志存储中心；所述用户已经通过身份识别。

4、根据权利要求3所述的一种用户兴趣偏好向量生成系统，其特征在于，所述服务平台是下面中的至少一种：即时通讯工具、网站业务平台、博客系统、移动通讯工具、论坛或邮件系统。

5、根据权利要求1所述的一种用户兴趣偏好向量生成系统，其特征在于：所述用户兴趣偏好向量至少包括用户使用服务平台的地理位置偏好、语言偏好、类别偏好、其浏览网页的语义偏好、产品品牌偏好。

6、一种用户兴趣偏好向量生成方法，其特征在于，该方法包括以下步骤：

(a)，所述主题语义提取模块对所述交互日志存储中心存储的一段时间内的用户点击交互日志进行分析，得到一个按语义强度排序的主题分类标识向量，并发送到所述兴趣偏好计算模块；

(b)，所述兴趣偏好计算模块对相应用户的静态登记信息、主题分类标识向量和历史兴趣偏好向量信息进行综合分析，得到相关用户的当前兴趣偏好向量。

7、根据权利要求6所述的一种用户兴趣偏好向量生成方法，其特征在于：步骤(a)之前还包括步骤(a0)，用户开始使用服务平台后，所述用户点击交互日志记录模块至少获取相应用户的主动IP信息、开始使用服务发生时间、持续时间、交互信息，并将上述信息存储到所述点击交互日志存储中心。

8、根据权利要求7所述的一种用户兴趣偏好向量生成方法，其特征在于：步骤(a)中，所述主题语义提取模块定时对所述交互日志存储中心存储的、一段时间内的用户点击交互日志进行分析。

9、根据权利要求7所述的一种用户兴趣偏好向量生成方法，其特征在于：步骤(a0)中，用户开始使用服务平台后，所述用户身份识别模块首先对用户进行身份识别。

10、根据权利要求6所述的一种用户兴趣偏好向量生成方法，其特征在于：步骤(b)后进一步包括步骤(b1)，所述兴趣偏好计算模块将用户的当前兴趣偏好向量存储到所述用户兴趣偏好向量库中。

## 一种用户兴趣偏好向量生成系统和方法

### 技术领域

本发明涉及网络通信，更具体地说，涉及一种用户兴趣偏好向量生成系统和方法。

### 背景技术

现阶段多数服务平台如 IM（即时通信）工具、论坛系统、mail 系统都会 在用户注册的时候收集用户的年龄、性别、学历、兴趣等静态信息，但是对于 用户在使用 sp 服务时候的浏览内容、点击交互内容等动态信息没有很好的收 集、分析和利用。用户的兴趣偏好尚处于概念阶段，没有一个定量的数据结构的 描述，更没有商业应用。业务提供商对用户推送的信息和服务都是没有针对 性的，而用户不感兴趣的内容常常会对用户造成干扰，从而降低了用户对网站、 服务终端等的满意度和忠诚度。

### 发明内容

本发明的目的在于提供一种用户兴趣偏好向量生成系统和方法，使得服务 平台可以根据用户兴趣偏好向用户提供个性化服务。

本发明的技术方案是：一种用户兴趣偏好向量生成系统，包括，用户登记 信息存储中心，其用于存储用户的静态属性；点击交互日志存储中心，其用于 记录用户一段时间内的点击交互日志；还包括：主题语义提取模块和兴趣偏好 计算模块；所述主题语义提取模块用于对所述点击交互日志存储中心中存储的 一段时间内的用户点击交互日志进行分析，得到一个按语义强度排序的主题分 类标识向量；所述兴趣偏好计算模块用于从所述用户登记信息存储中心获取用 户静态登记信息、从所述主题语义提取模块获取所述主题分类标识向量，并对

上述信息进行分析,得到相关用户的兴趣偏好向量。所述用户兴趣偏好向量至少包括用户使用服务平台的地理位置偏好、语言偏好、类别偏好、其浏览网页的语义偏好、产品品牌偏好。

进一步的,本发明系统还包括用户兴趣偏好向量库,其用于存储用户的兴趣偏好向量;相应的,所述兴趣偏好计算模块还同时从所述用户兴趣偏好向量库中获取用户的历史兴趣偏好向量,与所述静态登记信息和主题分类标识向量一起进行分析,以得到用户当前的兴趣偏好向量,并存储到所述用户兴趣偏好向量库中。

进一步的,本发明系统还包括:用户注册登录管理模块,其用于处理用户的注册注销、注册信息的修改、用户的身份认证,管理所述用户登记信息存储中心,并为用户身份识别模块提供用户数据;用户身份识别模块,其用于对用户进行身份识别;用户点击交互日志记录模块,其用于记录用户本次在服务平台上的行为,并存储到所述点击交互日志存储中心;所述用户已经通过身份识别。

本发明所述服务平台是下面中的至少一种:即时通讯工具、网站业务平台、博客系统、移动通讯工具、论坛或邮件系统。

本发明还提供了一种用户兴趣偏好向量生成方法,该方法包括以下步骤:

(a),所述主题语义提取模块对所述交互日志存储中心存储的一段时间内的用户点击交互日志进行分析,得到一个按语义强度排序的主题分类标识向量,并发送到所述兴趣偏好计算模块;

(b),所述兴趣偏好计算模块对相应用户的静态登记信息、主题分类标识向量和历史兴趣偏好向量信息进行综合分析,得到相关用户的当前兴趣偏好向量。

进一步的,步骤(a)之前还包括步骤(a0),用户开始使用服务平台后,所述用户点击交互日志记录模块至少获取相应用户的主动IP信息、开始使用服务发生时间、持续时间、交互信息,并将上述信息存储到所述点击交互日志存储中心。步骤(a)中,所述主题语义提取模块定时对所述交互日志存储中心存储的、一段时间内的用户点击交互日志进行分析。且,步骤(a0)中,用

户开始使用服务平台后,所述用户身份识别模块首先对用户进行身份识别。

进一步的,步骤(b)后进一步包括步骤(b1),所述兴趣偏好计算模块将用户的当前兴趣偏好向量存储到所述用户兴趣偏好向量库中。

本发明的有益效果在于:可通过用户的登记信息、历史兴趣偏好信息、浏览过的页面、聊天内容等得到用户的兴趣偏好向量。服务平台就可以根据用户的兴趣偏好向量有针对性、有选择性地推送信息或者提供服务,突出用户感兴趣的信息和服务,减少对用户的骚扰,以此提升用户对服务平台的满意度、忠诚度。

## 附图说明

图1是本发明一种用户兴趣偏好向量生成系统的结构图。

图2是本发明一种用户兴趣偏好向量生成方法的一个应用实例的流程图。

## 具体实施方式

下面根据附图和具体实施例对本发明作进一步阐述。

本发明中所述的服务平台可以是即时通讯工具(IM)、论坛、网站业务平台、博客系统、移动通讯工具、邮件系统等。所述的一种用户兴趣偏好向量生成系统设置在服务器端。其中,网站业务平台可以是订票网等以销售为主的网站或新闻类网站,移动通讯工具可以是带有上网功能的手机、PDA(掌上电脑)等无线便携机。

如图1所示,一种用户兴趣偏好向量生成系统包括用户注册登录管理模块1、用户登记信息存储中心2、用户身份识别模块3、用户点击交互日志记录模块4、点击交互日志存储中心5、主题语义提取模块6、兴趣偏好计算模块7和用户兴趣偏好向量库8。

其中,用户注册登录管理模块1负责处理登录到服务平台上的用户的注册、注销、注册信息的修改以及用户的身份认证等消息,并实时更新用户登记信息存储中心2内的信息。它是用户登记信息存储中心2的管理者,同时也为用户

身份识别模块 3 提供用户身份数据。它与服务平台、用户登记信息存储中心 2、用户身份识别模块 3 相连。

用户登记信息存储中心 2 用于存储用户的登记信息。这些信息可以是用户注册时候填写的，也可以是后来登录后修改的。可以包括用户的年龄、性别、学历、住址、联系方式和兴趣爱好等信息。用户登记信息存储中心 2 与用户注册登录管理模块 1、兴趣偏好计算模块 7 相连。

用户身份识别模块 3 主要负责未登录用户的身份识别，从用户注册登录管理模块下载能够确定用户身份的信息。对未登录用户可以通过 IP 来标识，以 IP 地址来判定是哪个用户。同时也负责所有用户的状态维护和用户静态登记信息的获取。它与服务平台、用户注册登录管理模块 1 和用户点击交互日志记录模块 4 相连。

用户点击交互日志记录模块 4 负责记录用户本次在服务平台上的行为，并将记录信息存储到点击交互日志存储中心 5。其记录的信息可以包括浏览论坛、网页的发生时间、持续时间、uri（统一资源标志符）、主动 IP 信息、浏览网页的点击交互内容等。对于 IM 系统，该模块还负责记录用户的聊天内容。它与用户身份识别模块 3 和点击交互日志存储中心 5 相连。

点击交互日志存储中心 5 用于存储来自用户点击交互日志记录模块 4 记录的信息。限于存储器的容量，存储中心一般只存储用户一定时期内，如三个月的记录。

主题语义提取模块 6 负责定时或按照设定时间对一段时间内用户的点击交互日志进行分析，计算出每次浏览或聊天的主题内容，并且把这些内容按照当前服务平台的语义分类体系进行按语义强度的排序，得到排名前  $n$  个的主题分类标识的向量。上述一段时间可以是一个月内、一周内或其它指定期限。主题语义提取模块 6 与点击交互日志存储中心 5、兴趣偏好计算模块 7 相连。

兴趣偏好计算模块 7 负责计算出相应用户的兴趣偏好向量，其通过对用户的静态登记信息、动态点击交互主题语义（即上述主题分类标识向量）、以及用户的历史兴趣偏好向量进行综合分析，以得到相应用户的新的兴趣偏好向量，并存储到用户兴趣偏好向量库 8。

用户兴趣偏好向量可以包括：用户浏览论坛、网页的地理位置偏好、语言偏好、服务类别偏好、信息类别偏好、用户浏览网页的语义偏好、产品品牌偏好等点击偏好信息。其中，地理位置偏好是指用户偏向于在什么地方上网，比如用户的历史兴趣偏好向量显示用户的地理位置偏好是北京，但是最近在上海，则服务平台会重定向用户到上海的服务器。

用户兴趣偏好向量库8负责存储用户兴趣偏好计算模块7生成的用户兴趣偏好向量。一般来说，其存储的用户兴趣偏好向量由最近计算出的兴趣偏好向量与衰减后的历史兴趣偏好向量相迭加而成。可以定义历史兴趣偏好向量衰减的半周期为：每周衰减为原来的  $1/2$ 。该向量库可随用户新的网上点击交互行为发生更新。

在一种实际应用中，服务平台或服务器中的业务相关度计算模块9会定时或按照设定时间，从用户兴趣偏好向量库8中获取用户的兴趣偏好向量，并进行业务相关度的计算，并将结果发送到差异化业务提供模块10。所述差异化业务提供模块10根据业务的相关程度为用户提供有针对性的服务，向用户推送其感兴趣的广告或页面等，并由用户点击交互日志记录模块4实时记录用户对新推送服务的动作。例如，业务相关度计算模块9获得用户A的兴趣偏好向量为 `interest_vector[{数码, 数码相机, 长焦}]`，于是在其业务库中搜索与长焦数码相机相关的广告信息或服务，并推送给用户。

下面以用户使用IM聊天为例介绍本发明一种用户兴趣偏好向量生成方法、及服务推送的应用实例的流程图：

S1，用户开始使用一个服务类网站或IM所提供的信息服务。

S2，用户身份识别模块3根据用户的登录情况或主动IP信息对用户身份进行识别；对于未登录用户按照用户的IP地址来标示用户。

S3，兴趣偏好计算模块7从用户登记信息存储中心2得到用户静态登记信息 `register-a`、从用户兴趣偏好向量库8得到用户历史兴趣偏好向量 `interest_vector-a`。服务器或服务平台上的业务相关度计算模块9也得到相关用户的兴趣偏好向量，并进行相关计算后输出到差异化业务提供模块10。

其中，用户注册信息 `register-a` 显示该用户为北京地区28岁的IT研发



人员。其历史偏好向量 `interest-vector-a` 显示用户最近关注数码相机的相关信息特别是长焦相机。

S4, 服务器或服务平台上的差异化业务提供模块 10 根据业务相关度计算模块 9 的输出, 向用户提供个性化、差异化的服务: 如向用户投放北京地区年轻人热衷的长焦数码相机广告。

S5, 用户点击交互日志记录模块 4 记录用户本次在服务平台上的动作, 包括主动 ip、发生时间、浏览 url 或者点击、交互的内容到用户点击交互日志记录模块 4, 并存储到点击交互日志存储中心 5。

S6, 主题语义提取模块 6 定时对点击交互日志存储中心 5 的内容进行分析, 提取出用户浏览信息的主题语义, 并将其发送到兴趣偏好计算模块 7。所述主题语义为一个按语义强度排序的主题分类标识向量。

例如, 主题语义提取模块 6 在定时器时间发生时候对用户的历史交互日志进行分析, 发现用户和朋友 B 聊天的时候主要是在聊宝马轿车, 其次聊了诺基亚 3310 手机, 则将 A 的本次浏览语义计算为 `interest-vector-a-new`{ {汽车, 轿车, 宝马}, {数码, 手机, 诺基亚, 诺基亚 3310} }。

S7, 兴趣偏好计算模块 7 根据其获得的用户的登记信息 `register-a`、历史兴趣偏好向量 `interest-vector-a` 和浏览信息主题语义 `interest-vector-a-new` 进行分析计算, 得到用户的当前兴趣偏好向量 `interest-vector`{ {数码, 数码相机, 长焦}, {汽车, 轿车, 宝马}, {数码, 手机, 诺基亚, 诺基亚 3310} }。

S8, 兴趣偏好计算模块 7 将用户的当前兴趣偏好向量存储到用户兴趣偏好向量库 8, 用户兴趣偏好向量库 8 进行更新。

S9, 业务相关度计算模块 9 从用户兴趣偏好向量库 8 中获得用户的最新兴趣偏好向量, 并进行相关计算后输出到差异化业务提供模块 10, 差异化业务提供模块 10 根据业务相关度计算模块 9 的输出向用户提供相关个性化、差异化的服务, 如宝马轿车或诺基亚手机的广告。

S10, 一次用户兴趣偏好向量生成及更加其提供个性化服务的过程结束。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的权利要求范围之内。

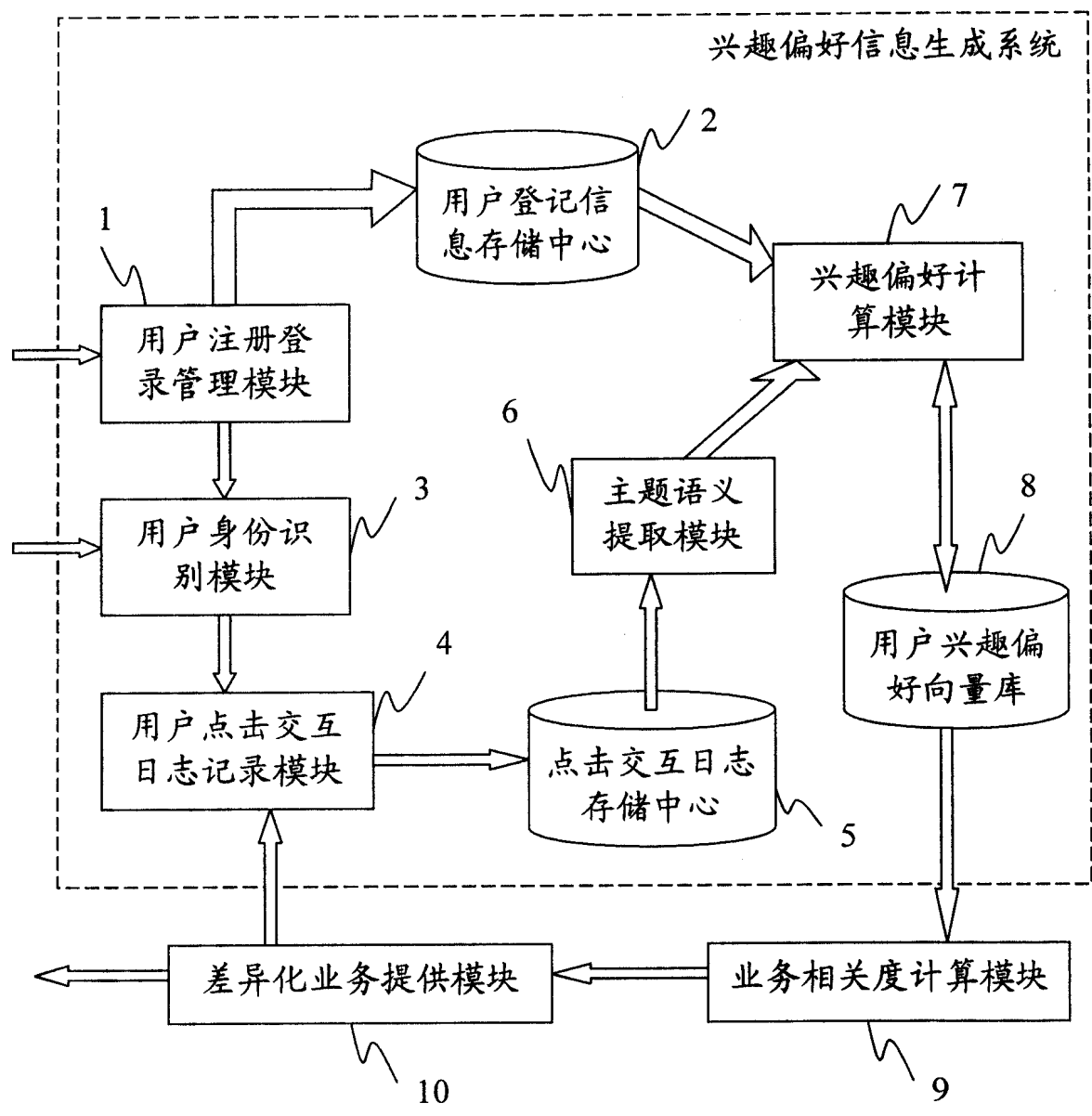


图 1

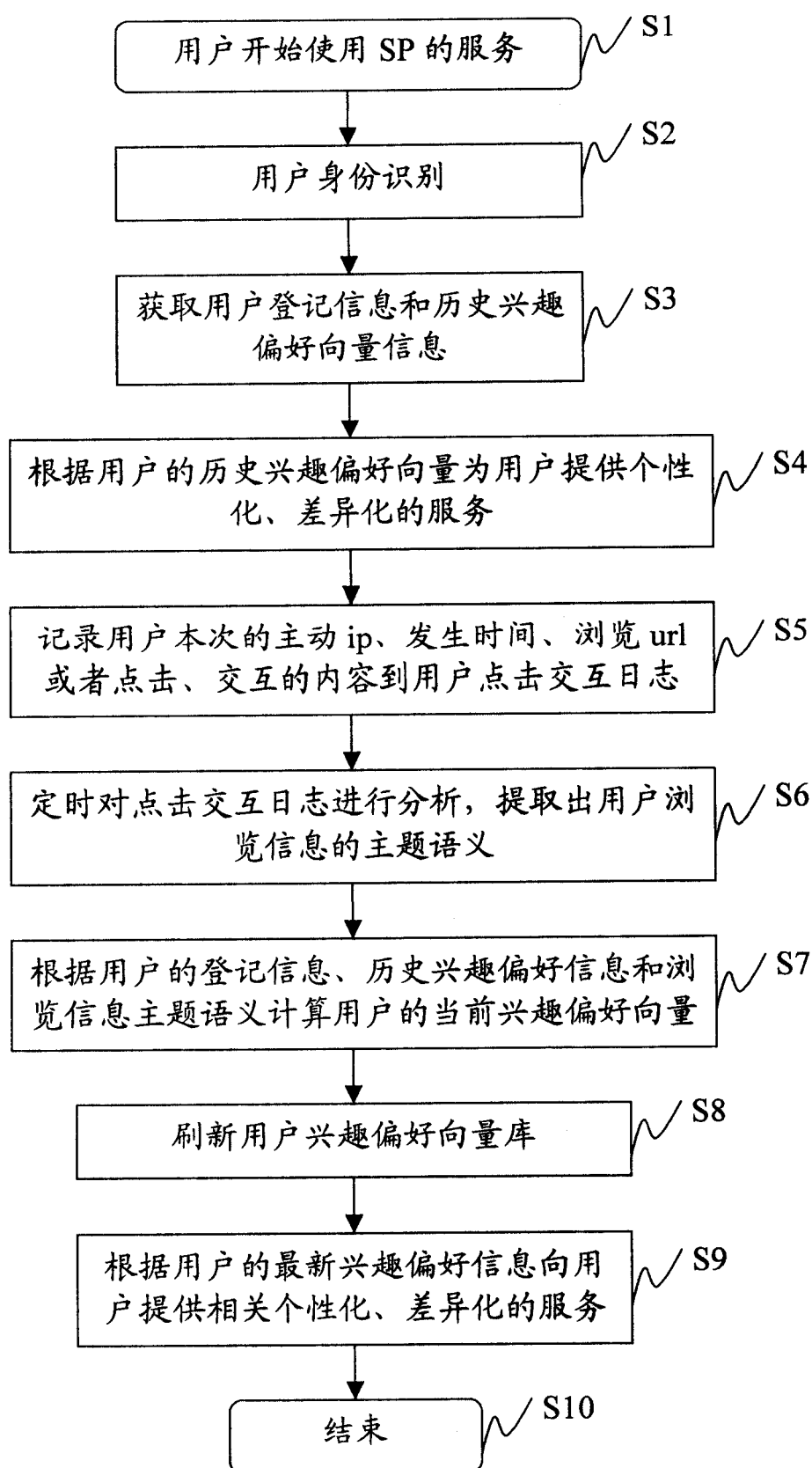


图 2