



(12)发明专利

(10)授权公告号 CN 106658564 B

(45)授权公告日 2019.10.25

(21)申请号 201611024350.7

(22)申请日 2016.11.17

(65)同一申请的已公布的文献号  
申请公布号 CN 106658564 A

(43)申请公布日 2017.05.10

(73)专利权人 广州杰赛科技股份有限公司  
地址 510310 广东省广州市海珠区新港中  
路381号31分箱

(72)发明人 杜翠凤

(74)专利代理机构 广州三环专利商标代理有限  
公司 44202

代理人 麦小婵 郝传鑫

(51)Int.Cl.

H04W 24/02(2009.01)

H04L 12/24(2006.01)

(56)对比文件

CN 105160547 A,2015.12.16,

CN 104244314 A,2014.12.24,

CN 102541886 A,2012.07.04,

CN 101841607 A,2010.09.22,

CN 102215504 A,2011.10.12,

审查员 杨雪

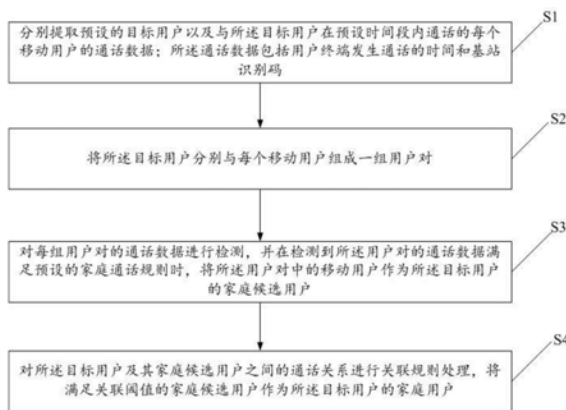
权利要求书3页 说明书8页 附图2页

(54)发明名称

一种家庭用户的识别方法及装置

(57)摘要

本发明公开了一种家庭用户的识别方法,包括:分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;将所述目标用户分别与每个移动用户组成一组用户对;对每组用户对中的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户。相应的,本发明还公开了一种家庭用户的识别装置。采用本发明实施例,能够快速、准确地识别家庭用户。



1. 一种家庭用户的识别方法,其特征在于,包括:

分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;

将所述目标用户分别与每个移动用户组成一组用户对;

对每组用户对中的通话数据进行检测,并在检测到所述用户对中的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;

对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户;

其中,所述关联阈值包括支持度阈值和置信度阈值;

所述对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户,具体包括:

获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系;

对获取的所有通话关系进行关联规则处理,获得每个交往对的支持度和置信度;其中,支持度为交往对中同时出现两个用户的概率,置信度为交往对中出现一个用户的情况下出现另一个用户的概率;

提取所述支持度满足所述支持度阈值,且置信度满足所述置信度阈值的交往对,并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。

2. 如权利要求1所述的家庭用户的识别方法,其特征在于,所述分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据,具体包括:

获取在预设时间段内与预设的目标用户通话的移动用户的话单;

去除所述话单为外省公司漫游话单、客服话单、公共服务性话单的移动用户;

分别提取所述目标用户和保留的每个移动用户在预设时间段内的通话数据。

3. 如权利要求1所述的家庭用户的识别方法,其特征在于,在所述将所述目标用户分别与每个移动用户组成一组用户对之前,还包括:

根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数;

去除所述交往指数低于预设交往阈值的移动用户,并将保留的移动用户添加到所述目标用户的交往圈中。

4. 如权利要求3所述的家庭用户的识别方法,其特征在于,所述根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数,具体包括:

根据所述通话数据,获取所述目标用户与每个移动用户的通话频率、通话次数和通话时长;

将所有移动用户按照通话次数的大小进行排序,并查询预设的通话次数排序得分表,获得每个移动用户所对应的通话次数得分;

将所有移动用户按照通话时长的大小进行排序,并查询预设的通话时长排序得分表,获得每个移动用户所对应的通话时长得分;

根据所述通话频率、所述通话次数得分和所述通话时长得分,采用交往指数算法,计算获得每个移动用户与所述目标用户的交往指数;

所述交往指数算法如下:

$$F=40*X+30*Y+30*Z;$$

其中,F为交往指数,X为通话频率,Y为通话次数得分,Z为通话时长得分。

5.如权利要求3所述的家庭用户的识别方法,其特征在于,所述家庭通话规则包括以下任意一种:

用户对中的目标用户和移动用户对其交往圈的总通话占比均大于40%,且通话总时长均大于50分钟;

用户对中的目标用户和移动用户的交往圈重合率达到20%,且对其交往圈的总通话占比均大于40%;

用户对中的目标用户和移动用户对其交往圈的源呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

用户对中的目标用户和移动用户对其交往圈的目标呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

用户对中的目标用户和移动用户的闲时通话基站重合率达到30%,且与其交往圈的通话次数均大于5次;

用户对中的目标用户和移动用户与其交往圈的通话次数大于15次,且周末通话基站重合率达到30%。

6.如权利要求5所述的家庭用户的识别方法,其特征在于,所述家庭用户的识别方法还包括:

根据获取的所有通话关系,采用频繁模式增长算法,从所述目标用户及其家庭用户中识别出家庭户主。

7.一种家庭用户的识别装置,其特征在于,包括:

数据提取模块,用于分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;

用户对组成模块,用于将所述目标用户分别与每个移动用户组成一组用户对;

检测模块,用于对每组用户对的通话数据进行检测,并在检测到所述用户对的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;以及,

家庭用户获取模块,用于对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户;

其中,所述关联阈值包括支持度阈值和置信度阈值;

所述家庭用户获取模块具体包括:

通话关系获取单元,用于获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系;

数据获取单元,用于对获取的所有通话关系进行关联规则处理,获得每个交往对的支持度和置信度;其中,支持度为交往对中同时出现两个用户的概率,置信度为交往对中出现一个用户的情况下出现另一个用户的概率;以及,

家庭用户获取单元,用于提取所述支持度满足所述支持度阈值,且置信度满足所述置信度阈值的交往对,并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。

8.如权利要求7所述的家庭用户的识别装置,其特征在于,所述家庭用户的识别装置还

包括：

户主识别模块，用于根据获取的所有通话关系，采用频繁模式增长算法，从所述目标用户及其家庭用户中识别出家庭户主。

## 一种家庭用户的识别方法及装置

### 技术领域

[0001] 本发明涉及移动通信技术领域,尤其涉及一种家庭用户的识别方法及装置。

### 背景技术

[0002] 随着移动通信的发展,家庭用户市场是近几年出现的一个新兴市场,家庭用户是对集团客户市场和个人客户市场的一个很好的补充。在网络带宽的不断升级,以及各类新兴增值业务快速发展的推动下,家庭用户已不满足于单一的语音、上网需求,对移动通信企业来说,用户的需求点转向了娱乐、生活应用类的服务需求,因此各大电信运营商陆续开始发展家庭业务,家庭用户市场蕴含着很多新的商机。但是,目前运营商仅仅关注用户的交往圈子,对于用户的家庭圈子的关注度较少,无法有效针对家庭用户推广家庭业务。

### 发明内容

[0003] 本发明实施例提出一种家庭用户的识别方法及装置,能够快速、准确地识别家庭用户。

[0004] 本发明实施例提供一种家庭用户的识别方法,包括:

[0005] 分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;

[0006] 将所述目标用户分别与每个移动用户组成一组用户对;

[0007] 对每组用户对中的通话数据进行检测,并在检测到所述用户对中的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;

[0008] 对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户。

[0009] 进一步地,所述分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据,具体包括:

[0010] 获取在预设时间段内与预设的目标用户通话的移动用户的话单;

[0011] 去除所述话单为外省公司漫游话单、客服话单、公共服务性话单的移动用户;

[0012] 分别提取所述目标用户和保留的每个移动用户在预设时间段内的通话数据。

[0013] 进一步地,在所述将所述目标用户分别与每个移动用户组成一组用户对之前,还包括:

[0014] 根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数;

[0015] 去除所述交往指数低于预设交往阈值的移动用户,并将保留的移动用户添加到所述目标用户的交往圈中。

[0016] 进一步地,所述根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数,具体包括:

[0017] 根据所述通话数据,获取所述目标用户与每个移动用户的通话频率、通话次数和通话时长;

- [0018] 将所有移动用户按照通话次数的大小进行排序,并查询预设的通话次数排序得分表,获得每个移动用户所对应的通话次数得分;
- [0019] 将所有移动用户按照通话时长的大小进行排序,并查询预设的通话时长排序得分表,获得每个移动用户所对应的通话时长得分;
- [0020] 根据所述通话频率、所述通话次数得分和所述通话时长得分,采用交往指数算法,计算获得每个移动用户与所述目标用户的交往指数;
- [0021] 所述交往指数算法如下:
- [0022]  $F=40*X+30*Y+30*Z$ ;
- [0023] 其中,F为交往指数,X为通话频率,Y为通话次数得分,Z为通话时长得分。
- [0024] 进一步地,所述家庭通话规则包括以下任意一种:
- [0025] 用户对中的目标用户和移动用户对其交往圈的总通话占比均大于40%,且通话总时长均大于50分钟;
- [0026] 用户对中的目标用户和移动用户的交往圈重合率达到20%,且对其交往圈的总通话占比均大于40%;
- [0027] 用户对中的目标用户和移动用户对其交往圈的源呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;
- [0028] 用户对中的目标用户和移动用户对其交往圈的目标呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;
- [0029] 用户对中的目标用户和移动用户的闲时通话基站重合率达到30%,且与其交往圈的通话次数均大于5次;
- [0030] 用户对中的目标用户和移动用户与其交往圈的通话次数大于15次,且周末通话基站重合率达到30%。
- [0031] 进一步地,所述关联阈值包括支持度阈值和置信度阈值;
- [0032] 所述对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户,具体包括:
- [0033] 获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系;
- [0034] 对获取的所有通话关系进行关联规则处理,获得每个交往对的支持度和置信度;
- [0035] 提取所述支持度满足所述支持度阈值,且置信度满足所述置信度阈值的交往对,并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。
- [0036] 进一步地,所述用户关系预测方法还包括:
- [0037] 根据获取的所有通话关系,采用频繁模式增长算法,从所述目标用户及其家庭用户中识别出家庭户主。
- [0038] 相应地,本发明实施例还提供一种家庭用户的识别装置,包括:
- [0039] 数据提取模块,用于分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;
- [0040] 用户对组成模块,用于将所述目标用户分别与每个移动用户组成一组用户对;
- [0041] 检测模块,用于对每组用户对的通话数据进行检测,并在检测到所述用户对的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭

候选用户;以及,

[0042] 家庭用户获取模块,用于对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户。

[0043] 进一步地,所述关联阈值包括支持度阈值和置信度阈值;

[0044] 所述家庭用户获取模块具体包括:

[0045] 通话关系获取单元,用于获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系;

[0046] 数据获取单元,用于对获取的所有通话关系进行关联规则处理,获得每个交往对的支持度和置信度;以及,

[0047] 家庭用户获取单元,用于提取所述支持度满足所述支持度阈值,且置信度满足所述置信度阈值的交往对,并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。

[0048] 进一步地,所述家庭用户的识别装置还包括:

[0049] 户主识别模块,用于根据获取的所有通话关系,采用频繁模式增长算法,从所述目标用户及其家庭用户中识别出家庭户主。

[0050] 实施本发明实施例,具有如下有益效果:

[0051] 本发明实施例提供的家庭用户的识别方法及装置,能够将目标用户分别与其通话的移动用户组成用户对,并根据其通话数据从移动用户中检测出目标用户的家庭候选用户,再结合关联规则对目标用户及其家庭候选用户之间的通话关系进行处理,以从家庭候选用户中识别出目标用户的家庭用户,实现家庭用户的快速、准确识别;采用频繁模式增长算法对目标用户及其家庭候选用户之间的通话关系进行处理,以识别出目标用户及其家庭用户所构成的家庭圈子中的户主,更有利于运营商在市场营销中抓住关键的人物。

## 附图说明

[0052] 图1是本发明提供的家庭用户的识别方法的一个实施例的流程示意图;

[0053] 图2是本发明提供的家庭用户的识别装置的一个实施例的结构示意图。

## 具体实施方式

[0054] 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、完整地描述,显然,所描述的实施例仅仅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作出创造性劳动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。

[0055] 参见图1,本发明提供的家庭用户的识别方法的一个实施例的流程示意图,包括:

[0056] S1、分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;

[0057] S2、将所述目标用户分别与每个移动用户组成一组用户对;

[0058] S3、对每组用户对的通话数据进行检测,并在检测到所述用户对的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;

[0059] S4、对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满

足关联阈值的家庭候选用户作为所述目标用户的家庭用户。

[0060] 需要说明的是,对目标用户的家庭用户进行识别需将与目标用户通话的每个移动用户分别与该目标用户进行匹配。先分别提取该目标用户及其每个移动用户的通话数据,其中,目标用户分别与每个移动用户组成一组用户对以进行检测匹配。当检测到一组用户对中的目标用户和移动用户相匹配,即其通话数据满足预设的家庭通话规则时,就保留该用户对中的移动用户,否则去除该用户对中的移动用户,其中,保留的移动用户为目标用户的家庭候选用户,即保留的移动用户有可能为目标用户的家庭用户。在筛选出家庭候选用户后,对目标用户及其家庭候选用户中的“一对一”交往对进行关联规则处理,计算每个交往对的关联值,并筛选出关联值大于预设的关联阈值的交往对,筛选出的交往对中的移动用户即为目标用户的家庭用户。本发明实施例利用运营商的移动用户动态、具有时效性的用户交往网络,采用家庭通话规则寻找出家庭候选用户,再结合关联规则的方法找到目标用户的家庭用户,而实现家庭用户的快速、准确地识别。

[0061] 进一步地,所述分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据,具体包括:

[0062] 获取在预设时间段内与预设的目标用户通话的移动用户的话单;

[0063] 去除所述话单为外省公司漫游话单、客服话单、公共服务性话单的移动用户;

[0064] 分别提取所述目标用户和保留的每个移动用户在预设时间段内的通话数据。

[0065] 需要说明的是,在提取通话数据时,可对与目标用户通话的移动用户进行初步筛选,即在于目标用户通话的话单中,对外省公司用户漫游来的话单、客服话单(如10086/10000)、公共服务性话单(如以400/800/600开头的的话单)进行剔除,以便目标用户的通话圈能够真正反映一个正常的通信用户的基本、规律、本地交往诉求。进而,对目标用户和保留的每个移动用户的通信数据进行提取,其中,通信数据为基于运营商的信息,这些信息很容易通过运营商获取,使得通信数据的获取便利。另外,本发明实施例考虑了用户交往网络的动态性,贴合运营商进行市场推广的特点。

[0066] 进一步地,在所述将所述目标用户分别与每个移动用户组成一组用户对之前,还包括:

[0067] 根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数;

[0068] 去除所述交往指数低于预设交往阈值的移动用户,并将保留的移动用户添加到所述目标用户的交往圈中。

[0069] 需要说明的是,在提取通信数据后,需根据通信数据对保留的移动用户进行第二次筛选,即分别计算每个移动与目标用户的交往指数,交往指数大的移动用户为目标用户的有效通信用户,从而将交往指数大于预设交往阈值的移动用户纳入目标用户的交往圈中。另外,在获取目标用户的交往圈的同时,采用相同方法获取目标用户交往圈中每个移动用户的交往圈,以便对移动用户进行下一步的筛选。

[0070] 进一步地,所述根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数,具体包括:

[0071] 根据所述通话数据,获取所述目标用户与每个移动用户的通话频率、通话次数和通话时长;

[0072] 将所有移动用户按照通话次数的大小进行排序,并查询预设的通话次数排序得分



表,获得每个移动用户所对应的通话次数得分;

[0073] 将所有移动用户按照通话时长的大小进行排序,并查询预设的通话时长排序得分表,获得每个移动用户所对应的通话时长得分;

[0074] 根据所述通话频率、所述通话次数得分和所述通话时长得分,采用交往指数算法,计算获得每个移动用户与所述目标用户的交往指数;

[0075] 所述交往指数算法如下:

[0076]  $F=40*X+30*Y+30*Z$ ;

[0077] 其中,F为交往指数,X为交往频率,Y为通话次数得分,Z为通话时长得分。

[0078] 需要说明的是,通话数据中可以看出目标用户与每个移动用户的通话频率、通话次数和通话时长,并分别针对每个移动用户进行通话次数打分和通话时长打分,进而根据这些分数计算出目标用户与每个移动用户之间的交往指数,交往指数大于预设交往阈值的移动用户才算是目标用户的有效交往用户,以纳入目标用户的交往圈,其中,预设交往阈值一般取50。

[0079] 进一步地,所述家庭通话规则包括以下任意一种:

[0080] 用户对中的目标用户和移动用户对其交往圈的总通话占比均大于40%,且通话总时长均大于50分钟;

[0081] 用户对中的目标用户和移动用户的交往圈重合率达到20%,且对其交往圈的总通话占比均大于40%;

[0082] 用户对中的目标用户和移动用户对其交往圈的源呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

[0083] 用户对中的目标用户和移动用户对其交往圈的目标呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

[0084] 用户对中的目标用户和移动用户的闲时通话基站重合率达到30%,且与其交往圈的通话次数均大于5次;

[0085] 用户对中的目标用户和移动用户与其交往圈的通话次数大于15次,且周末通话基站重合率达到30%。

[0086] 需要说明的是,用户对中的目标用户和移动用户对其交往圈的总通话占比均大于40%为,目标用户对其交往圈中所有用户的总通话占比大于40%,移动用户对其交往圈中所有用户的总通话占比也大于40%;用户对中的目标用户和移动用户对其交往圈的通话总时长均大于50分钟为,目标用户对其交往圈中所有用户的通话总时长大于50分钟,移动用户对其交往圈中所有用户的通话总时长也大于50分钟;用户对中的目标用户和移动用户的交往圈重合率达到20%为,目标用户和移动用户的交往圈重合的人数占目标用户交往圈的比例达到20%,占移动用户交往圈的比例也达到20%;用户对中的目标用户和移动用户对其交往圈的源呼出比例均大于45%为,目标用户呼叫其交往圈中所有用户的比例大于45%,移动用户呼叫其交往圈中所有用户的比例也大于45%;用户对中的目标用户和移动用户的闲时活动小区重合率达到30%为,目标用户和移动用户的用户终端闲时活动所重合的小区占目标用户的用户终端活动小区的比例达到30%,占移动用户的用户终端活动小区的比例也达到30%;用户对中的目标用户和移动用户对其交往圈的目标呼出比例均大于45%为,呼叫目标用户的用户中其交往圈中所有用户所占比例大于45%,呼叫移动用户的

用户中其交往圈中所有用户所占比例也大于45%；用户对中的目标用户和移动用户的闲时或周末通话基站重合率达到30%为，目标用户和移动用户的用户终端闲时或周末通话所重合的基站占目标用户的用户终端通话基站的比例达到30%，占移动用户的用户终端通话基站的比例也达到30%；用户对中的目标用户和移动用户与其交往圈的通话次数均大于5次为，目标用户与其交往圈中所有用户的通话总次数大于5次，移动用户与其交往圈中所有用户的通话总次数也大于5次。

[0087] 利用预设的家庭通话规则，通过决策树的分类能力对保留的移动用户进行第三次筛选。在目标用户与保留的移动用户所组成的每组用户对中，只要用户对中的目标用户和移动用户的通信数据满足家庭通话规则中的任意一种规则，则认为该用户对中的移动用户有可能是目标用户的家庭用户，保留该移动用户作为目标用户的家庭候选用户。

[0088] 进一步地，所述关联阈值包括支持度阈值和置信度阈值；

[0089] 所述对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理，将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户，具体包括：

[0090] 获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系；

[0091] 对获取的所有通话关系进行关联规则处理，获得每个交往对的支持度和置信度；

[0092] 提取所述支持度满足所述支持度阈值，且置信度满足所述置信度阈值的交往对，并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。

[0093] 需要说明的是，在获取目标用户的家庭候选用户后，需对家庭候选用户进行第四次筛选。从通话数据中可获取目标用户及其家庭候选用户之间的交往对，即一个用户呼叫另一个用户的通话关系，将获取的所有通话关系进行关联规则处理，计算每个交往对的支持度和置信度，其中，支持度为交往对中同时出现两个用户的概率，置信度为交往对中出现一个用户的情况下出现另一个用户的概率。提取大于预设支持度阈值且大于预设置信度阈值的交往对，提取的交往对中的所有用户即构成一个家庭圈子，该家庭圈子中的所有移动用户即为目标用户的家庭用户。

[0094] 进一步地，所述用户关系预测方法还包括：

[0095] 根据获取的所有通话关系，采用频繁模式增长算法，从所述目标用户及其家庭用户中识别出家庭户主。

[0096] 需要说明的是，在识别家庭用户的同时，本发明实施例还可采用频繁模式增长算法，即FP-growth算法对交往对的通话关系进行处理，高效发现频繁项集，例如，对于A、B、C、D用户，采用频繁模式增长算法发现A→B→C→D，且交往对AB、BA、AC、AD、DA、CB的支持度和置信度均大于预设阈值，则将A、B、C、D构成一个家庭圈子，而A为该家庭圈子的户主，其中，户主是在家庭圈子中具有主动性，在通信中处于主动地位的用户。本发明实施例计算复杂度不高，算法仅仅考虑用户交往网络的拓扑结构等因素，省去了传统考虑多种社会因素进行建模的麻烦，而且户主在网络节点的活跃度为营销推广的关键，同时，利用关联规则识别家庭用户，利用频繁模式增长算法识别户主具有一定的创新性。

[0097] 针对本发明实施例，通过联通内部用户的抽样验证来进行检验，本次验证主要提取了公司各部门员工的家庭用户数据，通过对业务支撑部门员工、市场部门员工、以及各分公司员工进行问卷调研后发现，抽样的家庭用户数据中58%为家庭成员，22%为密友，因此，抽样内部人员进行识别的准确率高达80%。

[0098] 本发明实施例提供的家庭用户的识别方法,能够将目标用户分别与其通话的移动用户组成用户对,并根据其通话数据从移动用户中检测出目标用户的家庭候选用户,再结合关联规则对目标用户及其家庭候选用户之间的通话关系进行处理,以从家庭候选用户中识别出目标用户的家庭用户,实现家庭用户的快速、准确识别;采用频繁模式增长算法对目标用户及其家庭候选用户之间的通话关系进行处理,以识别出目标用户及其家庭用户所构成的家庭圈子中的户主,更有利于运营商在市场营销中抓住关键的人物。

[0099] 相应的,本发明还提供一种家庭用户的识别装置,能够实现上述实施例中的家庭用户的识别方法的所有流程。

[0100] 参见图2,是本发明提供的家庭用户的识别装置的一个实施例的结构示意图,包括:

[0101] 数据提取模块1,用于分别提取预设的目标用户以及与所述目标用户在预设时间段内通话的每个移动用户的通话数据;所述通话数据包括用户终端发生通话的时间和基站识别码;

[0102] 用户对组成模块2,用于将所述目标用户分别与每个移动用户组成一组用户对;

[0103] 检测模块3,用于对每组用户对中的通话数据进行检测,并在检测到所述用户对中的通话数据满足预设的家庭通话规则时,将所述用户对中的移动用户作为所述目标用户的家庭候选用户;以及,

[0104] 家庭用户获取模块4,用于对所述目标用户及其家庭候选用户之间的通话关系进行关联规则处理,将满足关联阈值的家庭候选用户作为所述目标用户的家庭用户。

[0105] 进一步地,所述数据提取模块1具体包括:

[0106] 话单获取单元,用于获取在预设时间段内与预设的目标用户通话的移动用户的话单;

[0107] 去除单元,用于去除所述话单为外省公司漫游话单、客服话单、公共服务性话单的移动用户;以及,

[0108] 提取单元,用于分别提取所述目标用户和保留的每个移动用户在预设时间段内的通话数据。

[0109] 进一步地,所述家庭用户的识别装置还包括:

[0110] 交往指数计算模块,用于根据所述通话数据,分别计算每个移动用户与所述目标用户的交往指数;以及,

[0111] 交往圈获取模块,用于去除所述交往指数低于预设交往阈值的移动用户,并将保留的移动用户添加到所述目标用户的交往圈中。

[0112] 进一步地,所述交往指数计算模块具体包括:

[0113] 数据获取单元,用于根据所述通话数据,获取所述目标用户与每个移动用户的通话频率、通话次数和通话时长;

[0114] 第一得分获取单元,用于将所有移动用户按照通话次数的大小进行排序,并查询预设的通话次数排序得分表,获得每个移动用户所对应的通话次数得分;

[0115] 第二得分获取单元,用于将所有移动用户按照通话时长的大小进行排序,并查询预设的通话时长排序得分表,获得每个移动用户所对应的通话时长得分;以及,

[0116] 计算单元,用于根据所述通话频率、所述通话次数得分和所述通话时长得分,采用

交往指数算法,计算获得每个移动用户与所述目标用户的交往指数;

[0117] 所述交往指数算法如下:

[0118]  $F=40*X+30*Y+30*Z$ ;

[0119] 其中,F为交往指数,X为通话频率,Y为通话次数得分,Z为通话时长得分。

[0120] 进一步地,所述家庭通话规则包括以下任意一种:

[0121] 用户对中的目标用户和移动用户对其交往圈的总通话占比均大于40%,且通话总时长均大于50分钟;

[0122] 用户对中的目标用户和移动用户的交往圈重合率达到20%,且对其交往圈的总通话占比均大于40%;

[0123] 用户对中的目标用户和移动用户对其交往圈的源呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

[0124] 用户对中的目标用户和移动用户对其交往圈的目标呼出比例均大于45%,交往圈重合率达到20%,且闲时活动小区重合率达到30%;

[0125] 用户对中的目标用户和移动用户的闲时通话基站重合率达到30%,且与其交往圈的通话次数均大于5次;

[0126] 用户对中的目标用户和移动用户与其交往圈的通话次数大于15次,且周末通话基站重合率达到30%。

[0127] 进一步地,所述关联阈值包括支持度阈值和置信度阈值;

[0128] 所述家庭用户获取模块具体包括:

[0129] 通话关系获取单元,用于获取所述目标用户及其家庭候选用户之间的所有交往对的通话关系;

[0130] 数据获取单元,用于对获取的所有通话关系进行关联规则处理,获得每个交往对的支持度和置信度;以及,

[0131] 家庭用户获取单元,用于提取所述支持度满足所述支持度阈值,且置信度满足所述置信度阈值的交往对,并将提取的交往对中的家庭候选用户作为所述目标用户的家庭用户。

[0132] 进一步地,所述家庭用户的识别装置还包括:

[0133] 户主识别模块,用于根据获取的所有通话关系,采用频繁模式增长算法,从所述目标用户及其家庭用户中识别出家庭户主。

[0134] 本发明实施例提供的家庭用户的识别装置,能够将目标用户分别与其通话的移动用户组成用户对,并根据其通话数据从移动用户中检测出目标用户的家庭候选用户,再结合关联规则对目标用户及其家庭候选用户之间的通话关系进行处理,以从家庭候选用户中识别出目标用户的家庭用户,实现家庭用户的快速、准确识别;采用频繁模式增长算法对目标用户及其家庭候选用户之间的通话关系进行处理,以识别出目标用户及其家庭用户所构成的家庭圈子中的户主,更有利于运营商在市场营销中抓住关键的人物。

[0135] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

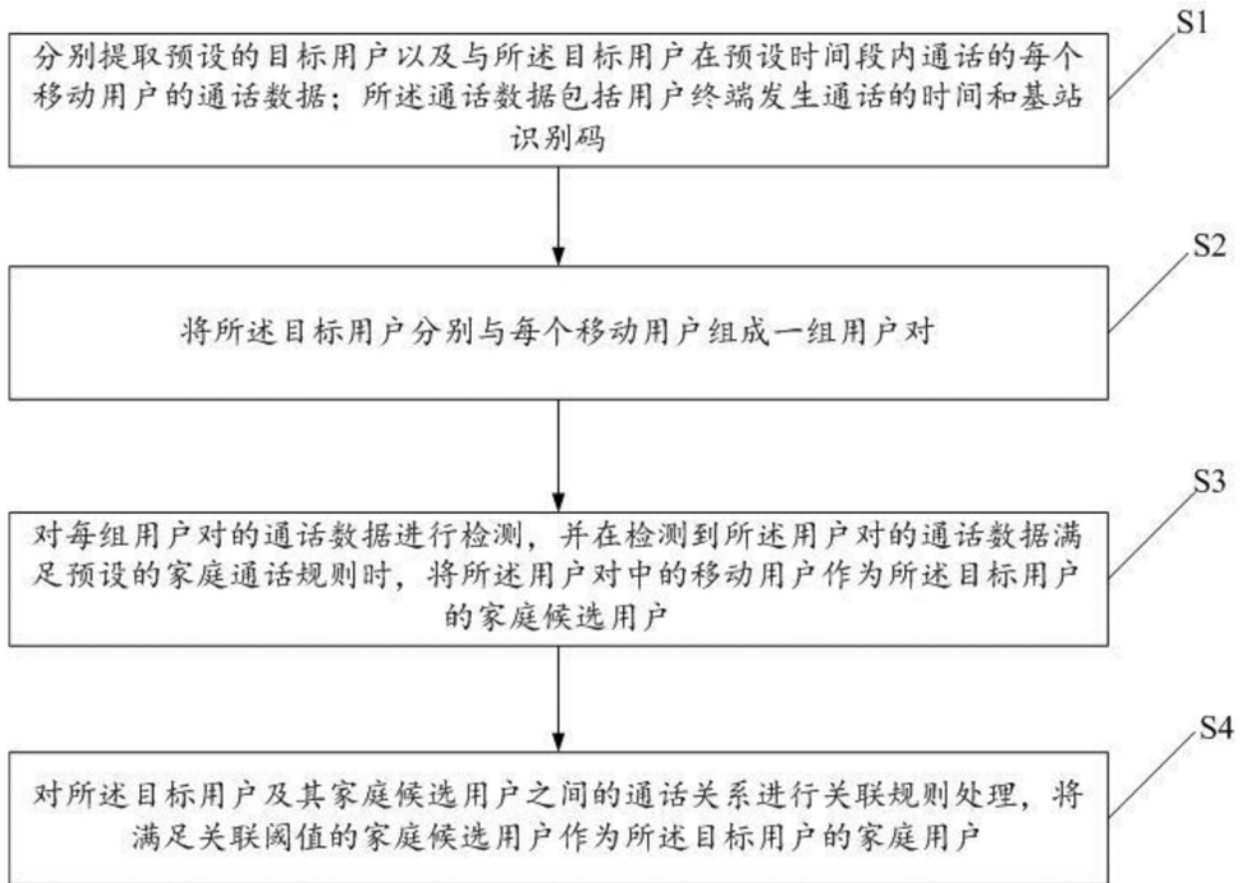


图1

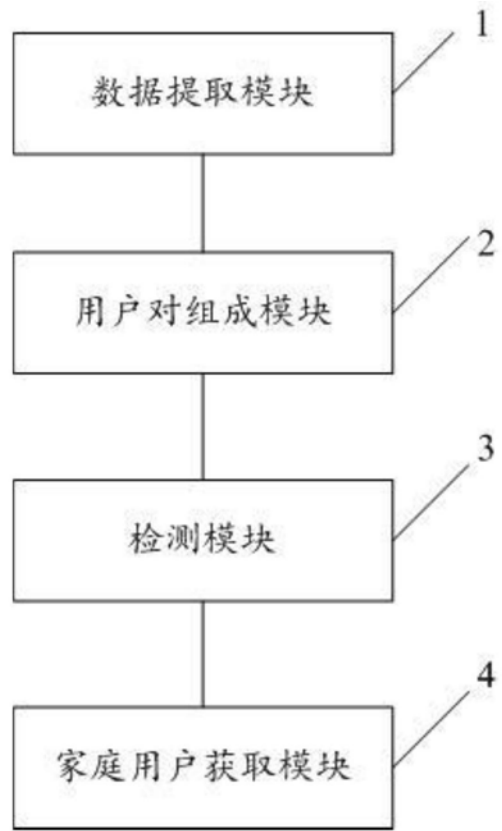


图2