



(12)发明专利

(10)授权公告号 CN 105721660 B

(45)授权公告日 2018.09.11

(21)申请号 201610076781.1

(56)对比文件

(22)申请日 2016.02.03

CN 104618554 A, 2015.05.13, 说明书第5页  
第5行-第8页第20行.

(65)同一申请的已公布的文献号

US 7552054 B1, 2009.06.23, 全文.

申请公布号 CN 105721660 A

CN 104717674 A, 2015.06.17, 全文.

(43)申请公布日 2016.06.29

审查员 贺永兴

(73)专利权人 北京光年无限科技有限公司

地址 100000 北京市石景山区石景山路3号  
玉泉大厦四层常青藤青年创业工作室  
193号

(72)发明人 杨钊

(74)专利代理机构 北京聿华联合知识产权代理  
有限公司 11611

代理人 张文娟 朱绘

(51)Int.Cl.

H04M 1/663(2006.01)

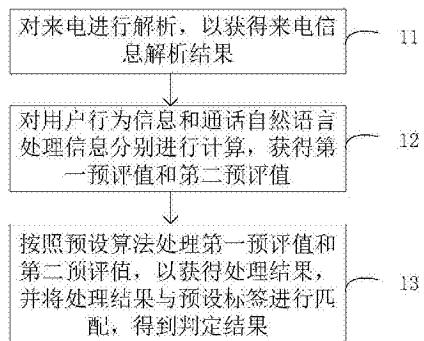
权利要求书2页 说明书7页 附图2页

(54)发明名称

骚扰电话识别方法及系统

(57)摘要

本发明提供一种骚扰电话识别方法及系统，其中，方法包括：来电解析步骤，对来电进行解析，以获得来电信息解析结果，其中，来电信息解析结果包括用户行为信息及通话自然语言处理信息；计算步骤，对用户行为信息和通话自然语言处理信息分别进行计算，获得第一预评估值和第二预评估值；判定步骤，按照预设算法处理第一预评估值和第二预评估值，以获得处理结果，并将处理结果与预设标签进行匹配，得到判定结果。此方法可通过对用户行为进行分析自动对来电进行判断来电是否为骚扰电话，从而减少用户处理骚扰电话的精力和时间，还可避免用户受到骚扰电话的骚扰、诈骗和恐吓等。



1. 一种骚扰电话识别方法,其特征在于,包括:

来电解析步骤,对来电进行解析,以获得来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息;

计算步骤,对所述用户行为信息和所述通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值;

判定步骤,按照预设算法处理所述第一预评值和所述第二预评值,以获得处理结果,并将所述处理结果与预设标签进行匹配,得到判定结果;

其中,判定步骤包括:

根据预设公式 $D=x*A+y*B$ 获得处理结果,其中,x为所述第一预评值的权重,y为所述第二预评值的权重,A为所述第一预评值,B为所述第二预评值。

2. 根据权利要求1所述的骚扰电话识别方法,其特征在于,计算步骤包括:

对用户行为信息赋予预设值,并对所述用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话进行处理,以获得第一预评值。

3. 根据权利要求1所述的骚扰电话识别方法,其特征在于,计算步骤还包括:

对用户通话内容进行监听,获得通话语音;

将所述通话语音转换成文本;

利用自然语言处理技术对所述文本处理,获得所述第二预评值。

4. 根据权利要求1-3任一所述的骚扰电话识别方法,其特征在于,在来电解析步骤之前,还包括:

预判步骤,对来电号码来进行预判,获得预判结果,所述预判结果包括骚扰电话和不确定两种;

若所述预判结果为不确定,则转所述来电解析步骤执行。

5. 根据权利要求4所述的骚扰电话识别方法,其特征在于,所述判定步骤在获得处理结果之后,还包括:

根据所述处理结果判断来电是骚扰电话、疑似骚扰电话或非骚扰电话,并给用户提供相应的提示信息;

若所述来电为骚扰电话,则将所述来电标记为骚扰电话并存储至数据库。

6. 一种骚扰电话识别系统,其特征在于,包括:

来电解析模块,用于接收来电信息并解析,得到来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息;

计算模块,用于对所述用户行为信息和所述通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值;

判定模块,用于按照预设算法处理所述第一预评值和所述第二预评值,以获得处理结果,并将所述处理结果与预设标签进行匹配,得到判定结果;

其中,判定模块具体用于:

根据预设公式 $D=x*A+y*B$ 获得处理结果,其中,x为所述第一预评值的权重,y为所述第二预评值的权重,A为所述第一预评值,B为所述第二预评值。

7. 根据权利要求6所述的骚扰电话识别系统,其特征在于,计算模块具体用于:

对所述用户行为信息赋予预设值,并对所述用户行为信息中的通话时长、通话频率、拨

号等待时长和主动拨号通话进行处理,以获得所述第一预评值。

8. 根据权利要求6所述的骚扰电话识别系统,其特征在于,计算模块还用于:
  - 对用户通话内容进行监听,获得通话语音;
  - 将所述通话语音转换成文本;
  - 利用自然语言处理技术对所述文本处理,获得所述第二预评值。

## 骚扰电话识别方法及系统

### 技术领域

[0001] 本发明涉及自然语言处理领域,尤其涉及一种骚扰电话识别方法及系统。

### 背景技术

[0002] 目前,在人们享受移动通信带来便利的同时,也常常遭受着骚扰电话的困扰,很多用户都会接到骗子或者推销人员等的骚扰电话,接收此类电话浪费时间且有上当风险。

[0003] 因而亟需一种对陌生来电号码自动进行判断的方法,判断该来电是否是骚扰电话,并且能够给予客户相应的提示信息,客户根据提示信息来决定是否接听电话。

### 发明内容

[0004] 本发明提供一种骚扰电话识别方法及系统,用以解决现有技术中没有一种方法可自动对来电进行判断以识别是否是骚扰电话的技术问题。

[0005] 本发明一方面提供一种骚扰电话识别方法,包括:

[0006] 来电解析步骤,对来电进行解析,以获得来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息;

[0007] 计算步骤,对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评估值和第二预评估值;

[0008] 判定步骤,按照预设算法处理第一预评估值和第二预评估值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。

[0009] 进一步的,计算步骤包括:

[0010] 对用户行为信息赋予预设值,并对用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话进行处理,以获得第一预评估值。

[0011] 进一步的,判定步骤包括:

[0012] 根据预设公式 $D = x*A + y*B$ 获得处理结果,其中,x为第一预评估值的权重,y为第二预评估值的权重,A为第一预评估值,B为第二预评估值。

[0013] 进一步的,计算步骤还包括:

[0014] 对用户通话内容进行监听,获得通话语音;

[0015] 将通话语音转换成文本;

[0016] 利用自然语言处理技术对文本处理,获得第二预评估值。

[0017] 进一步的,在来电解析步骤之前,还包括:

[0018] 预判步骤,对来电号码来进行预判,获得预判结果,预判结果包括骚扰电话和不确定两种;

[0019] 若预判结果为不确定,转来电解析步骤执行。

[0020] 进一步的,判定步骤在获得处理结果之后,还包括:

[0021] 根据处理结果判断来电是骚扰电话、疑似骚扰电话或非骚扰电话,并给用户提供相应的提示信息;

- [0022] 若来电为骚扰电话，则将来电标记为骚扰电话并存储至数据库。
- [0023] 本发明另一方面提供一种骚扰电话识别系统，包括：
  - [0024] 来电解析模块，用于接收来电信息并解析，得到来电信息解析结果，其中，来电信息解析结果包括用户行为信息及通话自然语言处理信息；
  - [0025] 计算模块，用于对用户行为信息和通话自然语言处理信息分别进行计算，获得第一预评估值和第二预评估值；
  - [0026] 判定模块，用于按照预设算法处理第一预评估值和第二预评估值，以获得处理结果，并将处理结果与预设标签进行匹配，得到判定结果。
  - [0027] 进一步的，计算模块具体用于：
    - [0028] 对用户行为信息赋予预设值，并对用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话进行处理，以获得第一预评估值。
    - [0029] 进一步的，判定模块具体用于：
      - [0030] 根据预设公式 $D = x*A + y*B$ 获得处理结果，其中，x为第一预评估值的权重，y为第二预评估值的权重，A为第一预评估值，B为第二预评估值。
    - [0031] 进一步的，计算模块还用于：
      - [0032] 对用户通话内容进行监听，获得通话语音；
      - [0033] 将通话语音转换成文本；
      - [0034] 利用自然语言处理技术对文本处理，获得第二预评估值。
  - [0035] 本发明提供的骚扰电话识别方法及系统，通过对接收的来电信息进行解析，获得用户行为信息和通话自然语言处理信息，并对用户行为信息和通话自然语言处理信息分别进行计算，获得第一预评估值和第二预评估值；按照预设算法处理第一预评估值和第二预评估值，以获得处理结果，并将处理结果与预设标签进行匹配，得到判定结果。此方法可通过对用户行为信息和通话自然语言处理信息进行判断，来确定来电是否为骚扰电话，从而减少用户处理骚扰电话的精力和时间，同时避免用户受到骚扰电话的骚扰、诈骗和恐吓等。

## 附图说明

- [0036] 在下文中将基于实施例并参考附图来对本发明进行更详细的描述。其中：
- [0037] 图1为根据本发明实施例一的骚扰电话识别方法的流程示意图；
- [0038] 图2为根据本发明实施例二的骚扰电话识别方法的流程示意图；
- [0039] 图3为根据本发明实施例三的骚扰电话识别系统的结构示意图；
- [0040] 图4为根据本发明实施例四的骚扰电话识别系统的结构示意图。
- [0041] 在附图中，相同的部件使用相同的附图标记。附图并未按照实际的比例绘制。

## 具体实施方式

- [0042] 下面将结合附图对本发明作进一步说明。
- [0043] 本发明提供一种骚扰电话识别方法及系统，用以解决现有技术中没有一种方法可自动对来电进行判断以识别是否是骚扰电话的技术问题。
- [0044] 实施例一
- [0045] 本实施例的执行主体为骚扰电话识别系统。

[0046] 图1为根据本发明实施例一的骚扰电话识别方法的流程示意图,如图1所示,本发明提供一种骚扰电话识别方法,包括:

[0047] 来电解析步骤11,对来电进行解析,以获得来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息。

[0048] 具体的,用户行为信息包括用户通话行为的信息,比如通话时长、通话频率、拨号等待时长和主动拨号通话等等,这些信息均可从手机的数据库中获得,或者从支持手机骚扰电话识别的服务器的数据库中获得,在此不做限定。通话自然语言处理信息包括用户通话内容,在用户授权同意监听通话内容的情况下,可利用手机自带软件对通话内容进行监听,并将通话内容存储至手机数据库中,或者支持手机骚扰电话识别的服务器的数据库中。

[0049] 计算步骤12,对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评估值和第二预评估值。

[0050] 具体的,利用用户行为信息进行计算,获得第一预评估值,利用通话自然语言处理信息进行计算,获得第二预评估值。

[0051] 判定步骤13,按照预设算法处理第一预评估值和第二预评估值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。

[0052] 具体的,将第一预评估值和第二预评估值按照预设算法进行处理,获得处理结果,该处理结果为一个数值,将该处理结果与预设标签进行匹配,即判断处理结果落在预设标签的哪个区间范围内,从而获知来电是骚扰电话、疑似骚扰电话或非骚扰电话,判定结果即为骚扰电话、疑似骚扰电话或非骚扰电话中的一项。预设标签具体可为数值区间,若预设标签设置有两个阈值,那么可获得三个数值区间,如第一个阈值为0.4,第二个阈值为0.8,则预设标签的三个数值区间分别为小于0.4的第一区间,大于0.4小于0.8的第二区间和大于0.8的第三区间,并且对于落在第一区间的处理结果,判定结果为骚扰电话,落在第二区间的处理结果,判定结果为疑似骚扰电话,落在第三区间的处理结果,判定结果为非骚扰电话。若为预设标签设置一个阈值,那么可获得2个数值区间,如若设置该阈值为0.4,那么此时预设标签的两个数值区间分别为小于0.4的第一区间和大于0.4的第二区间,落在第一区间的处理结果,判定结果为骚扰电话或疑似骚扰电话,落在第二区间的处理结果,判定结果为非骚扰电话。上述对于处理结果落在不同区间的获得不同的判定结果,可根据实际情况进行灵活变化,在此仅举例进行说明,并不做限定。

[0053] 本发明提供的骚扰电话识别方法,通过对接收的来电信息进行解析,获得用户行为信息和通话自然语言处理信息,并对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评估值和第二预评估值;按照预设算法处理第一预评估值和第二预评估值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。此方法可通过对用户行为信息和通话自然语言处理信息进行判断,即根据用户曾经对该来电行为信息的判断,来确定来电是否为骚扰电话,从而减少用户处理骚扰电话的精力和时间,同时避免用户受到骚扰电话的骚扰、诈骗和恐吓等。

[0054] 实施例二

[0055] 图2为根据本发明实施例二的骚扰电话识别方法的流程示意图,如图2所示,本发明提供一种骚扰电话识别方法,包括:

[0056] 预判步骤10,对来电号码来进行预判,获得预判结果,所述预判结果包括骚扰电话

和不确定两种；若所述预判结果为不确定，转来电解析步骤11执行。

[0057] 具体的，可通过在数据库中检索来电号码来进行预判，获得预判结果，预判结果包括骚扰电话和不确定两种，数据库专门用于存储骚扰电话号码，该数据库可存储在手机本地存储中，也可设置在服务器中，接收到来电之后，将该来电号码发送到数据库中，在数据库中进行检索，若数据库中已有对该来电号码为骚扰电话的记录，那么预判结果为骚扰电话，此时直接对用户给出该来电为骚扰电话的提示，若数据库中并未检索到该来电号码，那么预判结果为不确定，说明还需要进一步的对该来电号码进行判断，此时转来电解析步骤11执行。

[0058] 来电解析步骤11，对来电进行解析，以获得来电信息解析结果，其中，来电信息解析结果包括用户行为信息及通话自然语言处理信息。

[0059] 本步骤具体可参见实施例一中相应步骤的记载。

[0060] 计算步骤12，对用户行为信息和通话自然语言处理信息分别进行计算，获得第一预评值和第二预评值。

[0061] 进一步的，本步骤具体包括：

[0062] 步骤121，对用户行为信息赋予预设值，并对用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话进行处理，以获得第一预评值。

[0063] 具体的，预设值即为一个输入参数值，在对用户行为信息进行分析时，可将用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话与第一预评值的关系通过表达式来计算。对用户行为包括通话时长、通话频率、主动拨号通话和拨号等待时长等进行分析，如若通话时长超过5分钟，可认定为非骚扰电话，通话时长不足30秒，可认定为骚扰电话，通话时长在30秒和5分钟之间的，可认定为疑似骚扰电话，具体时间段可灵活设置，在此不做限定。若通话频率较高，如超过5次，可认定为非骚扰电话，若通过次数不超过2次，可认定为骚扰电话，通话频率在2次到5次之间的，可认定为疑似骚扰电话，若用户有主动拨号通话记录超过3次，可认定为非骚扰电话，若主动拨号通话记录只有1次，可认定为骚扰电话，主动拨号通话记录在1次到3次之间的，可认定为疑似骚扰电话。当然，还可以将通话时长、通话频率、主动拨号通话进行结合后进行判断，判断方法灵活多变，在此不再赘述，只要将用户行为包括通话时长、通话频率和主动拨号通话包含在内的判断方法，均在本发明保护范围之内。如为通话时长、通话频率、拨号等待时长和主动拨号通话设置各自的权重，该权重可体现用户行为信息中各信息对第一预评值的影响，此时的预设值即为该权重值，通过通话时长、通话频率、拨号等待时长和主动拨号通话与第一预评值的关系的表达式，计算获得第一预评值。

[0064] 进一步的，本步骤具体还包括：

[0065] 步骤122，对用户通话内容进行监听，获得通话语音；将通话语音转换成文本；利用自然语言处理技术对文本处理，获得第二预评值。

[0066] 自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究实现人与计算机之间通过自然语言进行有效沟通的各种理论和方法。由于它的研究涉及自然语言，即人们日常使用的语言，所以它与语言学的研究有密不可分的关系。它融语言学、计算机科学、数学、统计学于一体，目的就是构建能够有效的实现自然语言通信的体系。通过自然语言处理这一技术，能使机器具备理解自然语言的能力。

[0067] 具体的,本步骤能够执行的前提是用户对来电进行了接听,且同意骚扰电话识别系统对用户通话内容进行监听。骚扰电话识别系统将监听到的通话语音转换成文本,然后利用自然语言处理技术对文本内容进行分析,具体可通过大量的经过培训的语料和情景对文本进行处理,从而获得第二预评值。

[0068] 进一步的,如果用户对来电进行了接听,但是并没有同意骚扰电话识别系统对用户通话内容进行监听,那么从数据库中获取该来电的通话内容,从而将通话语音转换成文本,然后利用自然语言处理技术对文本内容进行分析,从而获得第二预评值,若数据库中也未存储该来电的通话内容,那么第二预评值为0。

[0069] 进一步的,步骤122还包括,将通话语音存储至数据库中,该数据库可为手机数据库,或者支持手机骚扰电话识别的服务器的数据库中。以便在用户不同意对通话内容进行监听时,从数据库中获取通话内容,从而计算获得第二预评值。

[0070] 判定步骤13,按照预设算法处理第一预评值和第二预评值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。

[0071] 进一步的,本步骤具体包括步骤131:根据预设公式 $D=x*A+y*B$ 获得处理结果,其中,x为第一预评值的权重,y为第二预评值的权重,A为第一预评值,B为第二预评值。

[0072] 具体的,D为来电的处理结果,同时考虑用户行为信息与用户通话内容的作用,两者各自的重要性可通过权重来体现,如若侧重用户行为信息对处理结果的影响,则可相应的增大第一预评值的权重x,若侧重通话内容对结果的影响,可相应的增大第二预评值的权重y,x与y的值可根据实际情况进行设置,如可将x与y均设置为小于1的值,且两者之和为1。当然,权重x与y还可以通过语义理解中的模型训练获得,以更加准确的对骚扰电话进行判断。

[0073] 进一步的,本步骤中在步骤131之后,还包括步骤132:根据处理结果判断来电是骚扰电话、疑似骚扰电话或非骚扰电话,并给用户提供相应的提示信息;若来电为骚扰电话,则将来电标记为骚扰电话并存储至数据库。

[0074] 具体的,若判定结果是骚扰电话,可提示用户“来电为骚扰电话”,让用户自己选择是否接听,也可直接自动挂断而不用等待用户处理,若判定结果是疑似骚扰电话,可提示用户“来电为疑似骚扰电话”,让用户自己选择是否接听,若判定结果是非骚扰电话,也可不用给出提示,直接等待用户接听。

[0075] 当然,若来电为骚扰电话,为了能够不断的更新数据库,可将来电标记为骚扰电话并存储到数据库中,下次再接收到该来电时,可直接在数据库中进行查找到,不需要再经过一系列的判定便可知该来电为骚扰电话。当然,还可以对非骚扰电话和疑似骚扰电话进行标记并存储至数据库中,在此不再赘述。

[0076] 本发明提供的骚扰电话识别方法,通过对接收的来电信息进行解析,获得用户行为信息和通话自然语言处理信息,并对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值;按照预设算法处理第一预评值和第二预评值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。此方法可通过对用户行为信息和通话自然语言处理信息进行判断,即根据用户曾经对该来电行为信息的判断,来确定来电是否为骚扰电话,从而减少用户处理骚扰电话的精力和时间,同时避免用户受到骚扰电话的骚扰、诈骗和恐吓等。

[0077] 实施例三

[0078] 本实施例为装置实施例,用以执行上述的骚扰电话识别方法。

[0079] 图3为根据本发明实施例三的骚扰电话识别系统的结构示意图,如图3所示,本发明提供一种骚扰电话识别系统,包括:来电解析模块21、计算模块22和判定模块23。

[0080] 其中,来电解析模块21,用于接收来电信息并解析,得到来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息;

[0081] 计算模块22,用于对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值;

[0082] 判定模块23,用于按照预设算法处理第一预评值和第二预评值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。

[0083] 本实施例是与方法实施例一对应的装置实施例,具体详情可参见实施例一中的描述,在此不再赘述。

[0084] 本发明提供的骚扰电话识别系统,通过来电解析模块21对接收的来电信息进行解析,获得用户行为信息和通话自然语言处理信息,并利用计算模块22对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值;最后通过判定模块23按照预设算法处理第一预评值和第二预评值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。此系统可通过对用户行为信息和通话自然语言处理信息进行判断,即根据用户曾经对该来电行为信息的判断,来确定来电是否为骚扰电话,从而减少用户处理骚扰电话的精力和时间,同时避免用户受到骚扰电话的骚扰、诈骗和恐吓等。

[0085] 实施例四

[0086] 本实施例是在实施例三的基础上进行的补充说明。

[0087] 图4为根据本发明实施例四的骚扰电话识别系统的结构示意图,如图4所示,本发明提供一种骚扰电话识别系统,包括:预判模块20,来电解析模块21、计算模块22和判定模块23。

[0088] 其中,预判模块20,用于对来电号码来进行预判,获得预判结果,预判结果包括骚扰电话和不确定两种;若预判结果为不确定,触发来电解析模块21。

[0089] 来电解析模块21,用于接收来电信息并解析,得到来电信息解析结果,其中,来电信息解析结果包括用户行为信息及通话自然语言处理信息。

[0090] 计算模块22,用于对用户行为信息和通话自然语言处理信息分别进行计算,获得第一预评值和第二预评值。

[0091] 进一步的,计算模块22具体用于:

[0092] 对用户行为信息赋予预设值,并对用户行为信息中的通话时长、通话频率、拨号等待时长和主动拨号通话进行处理,以获得第一预评值。

[0093] 进一步的,计算模块22具体还用于:

[0094] 对用户通话内容进行监听,获得通话语音;将通话语音转换成文本;利用自然语言处理技术对文本处理,获得第二预评值。

[0095] 判定模块23,用于按照预设算法处理第一预评值和第二预评值,以获得处理结果,并将处理结果与预设标签进行匹配,得到判定结果。

[0096] 进一步的,判定模块23具体用于:

[0097] 根据预设公式 $D=x*A+y*B$ 获得处理结果,其中,x为第一预评值的权重,y为第二预评值的权重,A为第一预评值,B为第二预评值。

[0098] 本实施例是与方法实施例二对应的装置实施例,具体详情可参见实施例二中的描述,在此不再赘述。

[0099] 虽然已经参考优选实施例对本发明进行了描述,但在不脱离本发明的范围的情况下,可以对其进行各种改进并且可以用等效物替换其中的部件。尤其是,只要不存在结构冲突,各个实施例中所提到的各项技术特征均可以任意方式组合起来。本发明并不局限于文中公开的特定实施例,而是包括落入权利要求的范围内的所有技术方案。

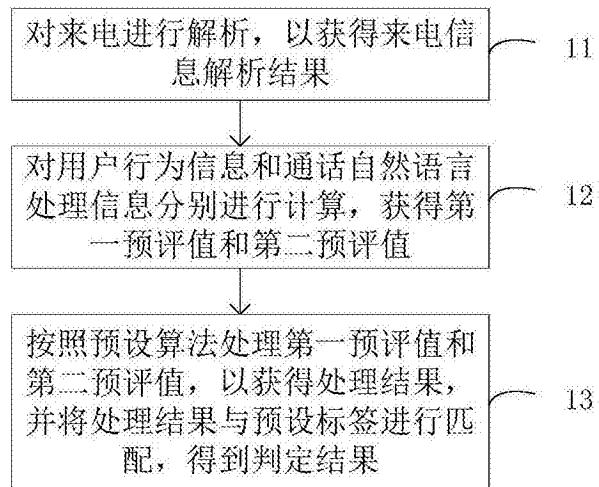


图1

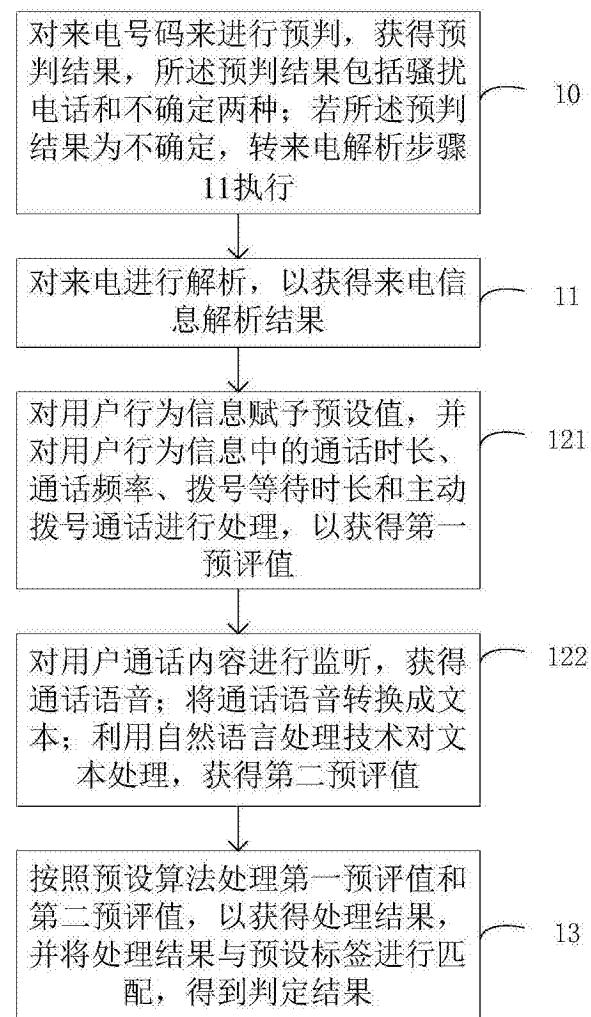


图2



图3



图4