



(12) 发明专利

(10) 授权公告号 CN 108920660 B

(45) 授权公告日 2020.11.20

(21) 申请号 201810723425.3

G06F 16/33 (2019.01)

(22) 申请日 2018.07.04

G06F 16/332 (2019.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 108920660 A

(56) 对比文件

CN 107102985 A, 2017.08.29

CN 106294314 A, 2017.01.04

(43) 申请公布日 2018.11.30

CN 107590195 A, 2018.01.16

(73) 专利权人 中国银行股份有限公司  
地址 100818 北京市西城区复兴门内大街1号

US 2017139899 A1, 2017.05.18

US 2011004610 A1, 2011.01.06

审查员 夏雪

(72) 发明人 宋雨

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 郟晨芳 王宝筠

(51) Int. Cl.

G06F 40/289 (2020.01)

G06F 40/216 (2020.01)

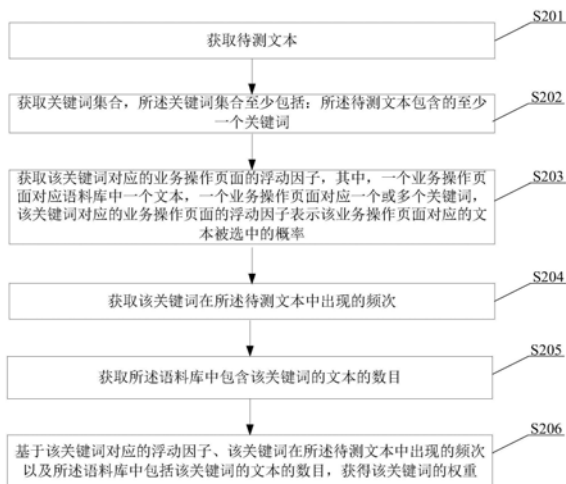
权利要求书2页 说明书9页 附图5页

(54) 发明名称

关键词权重获取方法、装置、电子设备及可读存储介质

(57) 摘要

本发明实施例提供了关键词权重获取方法、装置、电子设备及可读存储介质,由于考虑了关键词对应的业务操作页面的浮动因子,即相当于结合了各业务操作页面对应的文本被选中的概率,所以得到的关键词的权重才能够评估该关键词对于语料库包含的文件的重要程度;关键词的权重越大,表明该关键词对于语料库包含的文件越重要。基于关键词集合中每一关键词的权重,从语料库中的得到文件是用户意图查看的文本的概率就越大,即越准确。



1. 一种关键词权重获取方法,其特征在于,包括:
  - 获取待测文本;
  - 获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;
  - 针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;
  - 获取该关键词在所述待测文本中出现的频次;
  - 获取所述语料库中包含该关键词的文本的数目;
  - 基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包含该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。
2. 根据权利要求1所述关键词权重获取方法,其特征在于,还包括:
  - 基于所述关键词集合中每一关键词对应的权重,从所述语料库中获取目标文本。
3. 根据权利要求1或2所述关键词权重获取方法,其特征在于,所述获取待测文本包括:
  - 显示第一业务操作页面;
  - 响应于输入待测文本的操作指令,接收用户输入的所述待测文本。
4. 根据权利要求3所述关键词权重获取方法,其特征在于,所述关键词集合包括第一关键词和至少一个第二关键词,其中,所述第一关键词对应所述第一业务操作页面,所述针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,包括:
  - 针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率确定为该第二关键词对应的浮动因子;
  - 针对所述关键词集合中第一关键词,设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子;将所述第一值确定为所述第一关键词对应的浮动因子。
5. 根据权利要求4所述关键词权重获取方法,其特征在于,所述第一业务操作页面对应第一关键词,所述获取关键词集合包括:
  - 获取所述待测文本包括的所述至少一个关键词;
  - 若所述至少一个关键词不包括所述第一关键词,将所述第一关键词与所述至少一个关键词合并,得到所述关键词集合;
  - 所述获取该关键词在所述待测文本中出现的频次包括:
    - 设置所述第一关键词在所述待测文本中出现的频次为第二值。
6. 一种关键词权重获取装置,其特征在于,包括:
  - 第一获取模块,用于获取待测文本;
  - 第二获取模块,用于获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;
  - 第三获取模块,用于针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对

应的文本被选中的概率；

第四获取模块,用于针对所述关键词集合中任一关键词,获取该关键词在所述待测文本中出现的频次；

第五获取模块,用于针对所述关键词集合中任一关键词,获取所述语料库中包含该关键词的文本的数目；

第六获取模块,用于针对所述关键词集合中任一关键词,基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。

7. 根据权利要求6所述关键词权重获取装置,其特征在于,所述第一获取模块包括:

显示单元,用于显示第一业务操作页面；

接收单元,用于响应于输入待测文本的操作指令,接收用户输入的所述待测文本。

8. 根据权利要求7所述关键词权重获取装置,其特征在于,关键词集合包括第一关键词和至少一个第二关键词,其中,所述第一关键词对应所述第一业务操作页面,所述第三获取模块包括:

第一确定单元,用于针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率确定为该第二关键词对应的浮动因子；

第二确定单元,用于针对所述关键词集合中第一关键词,设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子;将所述第一值确定为所述第一关键词对应的浮动因子。

9. 一种电子设备,其特征在于,包括:

存储器,用于存储程序；

处理器,用于执行所述程序,所述程序具体用于:

获取待测文本；

获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词；

针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率；

获取该关键词在所述待测文本中出现的频次；

获取所述语料库中包含该关键词的文本的数目；

基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。

10. 一种可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现如权利要求1至5任一所述的关键词权重获取方法包含的各个步骤。

## 关键词权重获取方法、装置、电子设备及可读存储介质

### 技术领域

[0001] 本发明涉及加权技术领域,更具体的说,是涉及关键词权重获取方法、装置、电子设备及可读存储介质。

### 背景技术

[0002] TF-IDF (term frequency-inverse document frequency) 是一种用于信息检索与数据挖掘的常用加权技术,即利用字词的权重评估一个字词对于语料库包含的文件的重要程度。TF (Term Frequency) 是指字词在待测文本中出现词频, IDF (Inverse Document Frequency) 是指语料库中包含该字词的文档的数目的倒数,即逆文本频率指数。基于TF与IDF,可以利用待测文本中的字词的权重命中语料库中一个或多个文本。

[0003] 智能问答就是利用TF-IDF技术,基于用户输入的待测文本包含的字词,从语料库中确定出用户可能需要的文本,并展示给用户。例如,用户点击在线问答,并在展示的窗口中输入待测文本,例如,我要挂失;后台会基于“我要挂失”这一待测文本包含的“挂失”这一字词,从语料库中确定用户可能需要信用卡挂失文本或储蓄卡挂失文本,并将信用卡挂失文本或储蓄卡挂失文本展示给用户。

[0004] 一般情况下,待测文本都较短,导致各字词的TF均较小,例如,我要挂失中“挂失”这一字词的TF为1,使得利用TF-IDF技术,得到的字词的权重不准确,导致基于字词的权重从语料库中命中的文本不准确。

### 发明内容

[0005] 有鉴于此,本发明提供了一种关键词权重获取方法、装置、电子设备及可读存储介质。

[0006] 为实现上述目的,本发明提供如下技术方案:

[0007] 一种关键词权重获取方法,包括:

[0008] 获取待测文本;

[0009] 获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;

[0010] 针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;

[0011] 获取该关键词在所述待测文本中出现的频次;

[0012] 获取所述语料库中包含该关键词的文本的数目;

[0013] 基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。

- [0014] 其中,还包括:
- [0015] 基于所述关键词集合中每一关键词对应的权重,从所述语料库中获取目标文本。
- [0016] 其中,所述获取待测文本包括:
- [0017] 显示第一业务操作页面;
- [0018] 响应于输入待测文本的操作指令,接收用户输入的所述待测文本。
- [0019] 其中,所述关键词集合包括第一关键词和至少一个第二关键词,其中,所述第一关键词对应所述第一业务操作页面,所述针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,包括:
- [0020] 针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率确定为该第二关键词对应的浮动因子;
- [0021] 针对所述关键词集合中第一关键词,设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子;将所述第一值确定为所述第一关键词对应的浮动因子。
- [0022] 其中,所述第一业务操作页面对应第一关键词,所述获取关键词集合包括:
- [0023] 获取所述待测文本包括的所述至少一个关键词;
- [0024] 若所述至少一个关键词不包括所述第一关键词,将所述第一关键词与所述至少一个关键词合并,得到所述关键词集合;
- [0025] 所述获取该关键词在所述待测文本中出现的频次包括:
- [0026] 设置所述第一关键词在所述待测文本中出现的频次为第二值。
- [0027] 一种关键词权重获取装置,包括:
- [0028] 第一获取模块,用于获取待测文本;
- [0029] 第二获取模块,用于获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;
- [0030] 第三获取模块,用于针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;
- [0031] 第四获取模块,用于针对所述关键词集合中任一关键词,获取该关键词在所述待测文本中出现的频次;
- [0032] 第五获取模块,用于针对所述关键词集合中任一关键词,获取所述语料库中包含该关键词的文本的数目;
- [0033] 第六获取模块,用于针对所述关键词集合中任一关键词,基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包含该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。
- [0034] 其中,所述第一获取模块包括:
- [0035] 显示单元,用于显示第一业务操作页面;
- [0036] 接收单元,用于响应于输入待测文本的操作指令,接收用户输入的所述待测文本。
- [0037] 其中,关键词集合包括第一关键词和至少一个第二关键词,其中,所述第一关键词对应所述第一业务操作页面,所述第三获取模块包括:

[0038] 第一确定单元,用于针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率确定为该第二关键词对应的浮动因子;

[0039] 第二确定单元,用于针对所述关键词集合中第一关键词,设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子;将所述第一值确定为所述第一关键词对应的浮动因子。

[0040] 一种电子设备,包括:

[0041] 存储器,用于存储程序;

[0042] 处理器,用于执行所述程序,所述程序具体用于:

[0043] 获取待测文本;

[0044] 获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;

[0045] 针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;

[0046] 获取该关键词在所述待测文本中出现的频次;

[0047] 获取所述语料库中包含该关键词的文本的数目;

[0048] 基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。

[0049] 一种可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时,实现如上述任一所述的关键词权重获取方法包含的各个步骤。

[0050] 经由上述的技术方案可知,与现有技术相比,本发明公开了一种关键词权重获取方法,首先获取待测文本,并得到包含待测文本中至少一个关键词的关键词集合;针对关键词集合中每一关键词,得到该关键词对应的业务操作页面的浮动因子,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到关键词集合中每一个关键词的权重。由于考虑了关键词对应的业务操作页面的浮动因子,即相当于结合了各业务操作页面对应的文本被选中的概率,所以得到的关键词的权重才能够评估该关键词对于语料库包含的文件的重要程度;关键词的权重越大,表明该关键词对于语料库包含的文件越重要。基于关键词集合中每一关键词的权重,从语料库中的得到文件是用户意图查看的文本的概率就越大,即越准确。

## 附图说明

[0051] 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简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的实施例,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,还可以根据提供的附图获得其他的附图。

- [0052] 图1a至图1b为本发明实施例提供的智能问答的一种表现形式示意图；
- [0053] 图2为本发明实施例提供了关键词权重获取方法的一种实现方式的流程图；
- [0054] 图3为本发明实施例提供的关键词权重获取方法的另一种实现方式的流程图；
- [0055] 图4为本发明实施例提供的关键词权重获取装置的一种实现方式的结构图；
- [0056] 图5为本发明实施例提供的电子设备的一种实现方式的结构图。

## 具体实施方式

[0057] 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做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。

[0058] 本发明实施例提供的关键词权重获取方法，可以应用于客户端，客户端可以为网页客户端或应用程序客户端。

[0059] 客户端中经常会出现智能问答的情况，如图1a至图1b所示，为本发明实施例提供的智能问答的一种表现形式示意图。

[0060] 图1a中展示了客户端的某个业务操作页面以及在线问答提示框11。用户点击在线问答提示框11，即可进入图1b所示的智能问答窗口12。用户可以在图1b所示的智能问答窗口12输入待测文本。

[0061] 在一可选实施例中，用户还可以输入语音，客户端再将语音转换成待测文本。

[0062] 现有技术中，用户输入待测文本后，客户端可以基于待测文本中的关键词的权重，从语料库中找到相应的文本，并展示给用户，例如，用户输入“我要挂失”，客户端可以找到信用卡挂失文本和储蓄卡挂失文本，并提供给用户。由于用户输入的待测文本都比较短，导致待测文本包含的关键词在待测文本中出现的TF频次一般都较小，甚至都为1。通过TF无法体现关键词在待测文本中的重要性，使得基于TF和IDF得到的关键词的权重准确性交差，导致基于关键词的权重从语料库中的到的文本，不是用户所需要的文本。

[0063] 为了解决上述问题，本发明实施例提供了关键词权重获取方法，如图2所示，为本发明实施例提供了关键词权重获取方法的一种实现方式的流程图，该方法包括：

[0064] 步骤S201：获取待测文本。

[0065] 在一可选实施例中，用户可以输入待测文本，在另一可选实施例中，用户可以输入语音信号，客户端将语音信号转换成待测文本。

[0066] 步骤S202：获取关键词集合，所述关键词集合至少包括：所述待测文本包含的至少一个关键词。

[0067] 本发明实施例中预先为各业务操作页面分别配置一个或多个关键词，例如，储蓄卡操作页面分配的关键词为储蓄卡，信用卡操作页面配置的关键词为信用卡。储蓄卡转账页面配置的关键词为储蓄卡和转账。信用卡挂失页面配置的关键词为信用卡和挂失。

[0068] 待测文本包含的至少一个关键词中，每一关键词均对应一个业务操作页面。

[0069] 针对所述关键词集合中任一关键词执行步骤S203至步骤S206，以得到所述关键词集合中每一关键词对应的权重。

[0070] 步骤S203：获取该关键词对应的业务操作页面的浮动因子，其中，一个业务操作页

面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,该关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率。

[0071] 本发明实施例中,业务操作页面与其对应的语料库中文本的关系如下:

[0072] 客户端展示业务操作页面的过程,即为加载语料库中与该业务操作页面相应的文本的过程。

[0073] 假设,当前时间为2018年6月16日,可以统计在2018年1月1日至2018年6月1日期间,该客户端包含的各业务操作页面对应的文本被选中的次数。以得到在2018年1月1日至2018年6月1日期间,该客户端包含的各业务操作页面对应的文本被选中的概率。

[0074] 假设客户端包括3个业务操作页面,分别为业务操作页面1、业务操作页面2以及业务操作页面3,其中,业务操作页面1对应关键词1和关键词2,业务操作页面2对应关键词3,业务操作页面3对应关键词4和关键词5。

[0075] 在2018年1月1日至2018年6月1日期间,用户针对业务操作页面1进行操作的次数,即选中业务操作页面1对应的文本的次数为30次;用户针对业务操作页面2进行操作的次数,即选中业务操作页面2对应的文本的次数为50次;针对业务操作页面3进行操作的次数,即选中业务操作页面3对应的文本的次数为20次。

[0076] 那么关键词1和关键词2对应的浮动因子可以为 $30/(30+50+20)=0.3$ ;同理,关键词3对应的浮动因子可以为 $50/(30+50+20)=0.5$ ;关键词4和关键词5应的浮动因子可以为 $20/(30+50+20)=0.2$ 。

[0077] 步骤S204:获取该关键词在所述待测文本中出现的频次。

[0078] 假设待测文本为:我要挂失信用卡,那么待测文本包括的关键词为:挂失、信用卡;其中,信用卡在待测文本中出现的频次为1,挂失在待测文本中出现的频次为1。

[0079] 步骤S205:获取所述语料库中包含该关键词的文本的数目。

[0080] 假设语料库中包括:储蓄卡挂失文本,信用卡挂失文本,信用卡办理文文本,储蓄卡办理文本;其中,包含信用卡这一关键词的文本数目为2;包括挂失这一关键词的文本数目为2。

[0081] 步骤S203至步骤S205没有先后执行顺序。

[0082] 步骤S206:基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重。

[0083] 在一可选实施例中,一个关键词的权重=浮动因子\*TF\*IDF。假设一个关键词的浮动因子为0.3,该关键词在待测文本中出现的频次为5,语料库中包含该关键词的文本数目为10,则该关键词的权重= $0.3*5*1/10=0.15$ 。

[0084] 本发明实施例提供的关键词权重获取方法,首先获取待测文本,并得到包含待测文本中至少一个关键词的关键词集合;针对关键词集合中每一关键词,得到该关键词对应的业务操作页面的浮动因子,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到关键词集合中每一个关键词的权重。由于考虑了关键词对应的业务操作页面的浮动因子,即相当于结合了各业务操作页面对应的文本被选中的概率,所以得到的关键词的权重才能够评估该关键词对于语料库包含的文件的重要程度;关键词的权重越大,表明该关键



词对于语料库包含的文件越重要。基于关键词集合中每一关键词的权重,从语料库中的得到文件是用户意图查看的文本的概率就越大,即越准确。

[0085] 在一可选实施例中,上述关键词权重获取方法还包括:

[0086] 基于所述关键词集合中每一关键词对应的权重,从所述语料库中获取目标文本。

[0087] 具体的,将关键词集合包含的各关键词,依据各关键词分别对应的权重,进行降序排序;

[0088] 从语料库中,获取包括所述关键词集合中所有关键词的至少一个第一文本,展示所述至少一个第一文本;

[0089] 若语料库不包括所述第一文本,从关键词集合中去除最后M位关键词,得到第一关键词集合;M为大于或等于1的正整数;

[0090] 从语料库中,获取包含第一关键词集合中所有关键词的至少一个第二文本,展示所述至少一个第二文本;

[0091] 若语料库中不包括所述第二文本,从第一关键词集合中去除最后N位关键词,得到第二关键词集合;N为大于或等于1的正整数。

[0092] 从语料库中,获取包含第二关键词集合中所有关键词的至少一个第三文本,展示所述至少一个第三文本。

[0093] 本发明实施例中的目标文本可以包括:至少一个第一文本,和/或,至少一个第二文本,和/或,至少一个第三文本。

[0094] 如图3所示,为本发明实施例提供的关键词权重获取方法的另一种实现方式的流程图,该方法包括:

[0095] 步骤S301:显示第一业务操作页面,第一业务操作页面对应第一关键词。

[0096] 仍以图1a和图1b所示,图1a显示的第一业务操作页面为信用卡操作页面,则第一业务操作页面对应的第一关键词可以为信用卡。

[0097] 步骤S302:响应于输入待测文本的操作指令,接收用户输入的所述待测文本。

[0098] 假设,如图1b所示,待测文本为:我要挂失。

[0099] 步骤S303:获取关键词集合,所述关键词集合至少包括:第一关键词以及所述待测文本包含的至少一个第二关键词。

[0100] 在一可选实施例中,待测文本可能包括第一关键词,此时,所述至少一个第二关键词不包括第一关键词;在一可选实施例中,待测文本可能不包括第一关键词,此时,需要将所述第一关键词与所述至少一个第二关键词合并,得到所述关键词集合。且,由于第一关键词在待测文本中出现的频次为0,因此需要设置所述第一关键词在所述待测文本中出现的频次为第二值,第二值不等于0。第二值可以基于实际情况而定,例如,为1,或2,或3,...

[0101] 仍以图1a和图1b为例,则待测文本不包括第一关键词“信用卡”。待测文本仅包括关键词“挂失”,因此,得到的关键词集合包括:信用卡、挂失。

[0102] 步骤S304:设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子。

[0103] 步骤S305:将所述第一值确定为所述第一关键词对应的浮动因子。

[0104] 由于用户是在信用卡对应的业务操作页面进入智能问答的,因此,很大概率上,用户是需要针对信用卡进行操作。因此,设置的第一值较大,例如,大于任一第二关键词对应

的浮动因子。

[0105] 步骤S306:针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率作为该第二关键词对应的浮动因子,以得到所述至少一个第二关键词分别对应的浮动因子。

[0106] 假设,当前时间为2018年6月16日,预设时间段为2018年1月1日至2018年6月1日,假设客户端包括3个业务操作页面,分别为业务操作页面1、业务操作页面2以及业务操作页面3,其中,业务操作页面1对应关键词1和关键词2,业务操作页面2对应关键词3,业务操作页面3对应关键词4和关键词5。即至少一个第二关键词包括:关键词1、关键词2、关键词3、关键词4和关键词5。

[0107] 假设在2018年1月1日至2018年6月1日期间,用户针对业务操作页面1进行操作的次数,即选中业务操作页面1对应的文本的次数为30次;用户针对业务操作页面2进行操作的次数,即选中业务操作页面2对应的文本的次数为50次;针对业务操作页面3进行操作的次数,即选中业务操作页面3对应的文本的次数为20次。

[0108] 那么关键词1和关键词2对应的浮动因子可以为 $30/(30+50+20)=0.3$ ;同理,关键词3对应的浮动因子可以为 $50/(30+50+20)=0.5$ ;关键词4和关键词5应的浮动因子可以为 $20/(30+50+20)=0.2$ 。

[0109] 步骤S307:获取关键词集合中每一关键词在所述待测文本中出现的频次。

[0110] 步骤S308:获取所述语料库中包含关键词集合中每一关键词的文本的数目。

[0111] 步骤S304至步骤S308没有先后顺序。

[0112] 步骤S309:针对关键词集合中每一关键词,基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重,以得到所述关键词集合中每一关键词对应的权重。

[0113] 本发明实施例提供的关键词权重获取方法,考虑了进行智能问答时对应的第一业务操作页面,即结合了进行智能问答时的业务场景,使得基于关键词集合中每一关键词的权重,从语料库中得到的文本是用户意图查看的文本的概率就越大,即越准确。

[0114] 上述本发明公开的实施例中详细描述了方法,对于本发明的方法可采用多种形式的装置实现,因此本发明还公开了一种装置,下面给出具体的实施例进行详细说明。

[0115] 如图4所示,为本发明实施例提供的关键词权重获取装置的一种实现方式的结构图,该装置包括:

[0116] 第一获取模块41,用于获取待测文本;

[0117] 第二获取模块42,用于获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;

[0118] 第三获取模块43,用于针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;

[0119] 第四获取模块44,用于针对所述关键词集合中任一关键词,获取该关键词在所述待测文本中出现的频次;

- [0120] 第五获取模块45,用于针对所述关键词集合中任一关键词,获取所述语料库中包含该关键词的文本的数目;
- [0121] 第六获取模块46,用于针对所述关键词集合中任一关键词,基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。
- [0122] 可选的,还包括:
- [0123] 第七获取模块,用于基于所述关键词集合中每一关键词对应的权重,从所述语料库中获取目标文本。
- [0124] 可选的,所述第一获取模块包括:
- [0125] 显示单元,用于显示第一业务操作页面;
- [0126] 接收单元,用于响应于输入待测文本的操作指令,接收用户输入的所述待测文本。
- [0127] 可选的,关键词集合包括第一关键词和至少一个第二关键词,其中,所述第一关键词对应所述第一业务操作页面,所述第三获取模块包括:
- [0128] 第一确定单元,用于针对于所述关键词集合中任一第二关键词,将该第二关键词相应的业务操作页面对应的文本在预设时间内被选中的概率确定为该第二关键词对应的浮动因子;
- [0129] 第二确定单元,用于针对所述关键词集合中第一关键词,设置所述第一业务操作页面对应的文本被选中的概率为第一值,所述第一值大于或等于任一所述第二关键词对应的浮动因子;将所述第一值确定为所述第一关键词对应的浮动因子。
- [0130] 可选的,所述第一业务操作页面对应第一关键词,第二获取模块包括:
- [0131] 第一获取单元,用于获取所述待测文本包括的所述至少一个关键词;
- [0132] 第二获取单元,用于若所述至少一个关键词不包括所述第一关键词,将所述第一关键词与所述至少一个关键词合并,得到所述关键词集合;
- [0133] 所述第四获取模块包括:
- [0134] 设置单元,用于设置所述第一关键词在所述待测文本中出现的频次为第二值。
- [0135] 如图5所示,为本发明实施例提供的电子设备的一种实现方式的结构图,该电子设备包括:
- [0136] 存储器51,用于存储程序;
- [0137] 处理器52,用于执行所述程序,所述程序具体用于:
- [0138] 获取待测文本;
- [0139] 获取关键词集合,所述关键词集合至少包括:所述待测文本包含的至少一个关键词;
- [0140] 针对所述关键词集合中任一关键词,获取该关键词对应的业务操作页面的浮动因子,其中,一个业务操作页面对应语料库中一个文本,一个业务操作页面对应一个或多个关键词,一个关键词对应的业务操作页面的浮动因子表示该业务操作页面对应的文本被选中的概率;
- [0141] 获取该关键词在所述待测文本中出现的频次;
- [0142] 获取所述语料库中包含该关键词的文本的数目;
- [0143] 基于该关键词对应的浮动因子、该关键词在所述待测文本中出现的频次以及所述

语料库中包括该关键词的文本的数目,获得该关键词的权重,以得到所述关键词集合中每一关键词对应的权重。

[0144] 存储器51可能包含高速RAM存储器,也可能还包括非易失性存储器(non-volatile memory),例如至少一个磁盘存储器。

[0145] 处理器52可能是一个中央处理器CPU,或者是特定集成电路ASIC

[0146] (Application Specific Integrated Circuit),或者是被配置成实施本发明实施例的一个或多个集成电路。

[0147] 可选的,电子设备还可以包括通信总线53以及通信接口54,其中,存储器51、处理器52、通信接口54、通过通信总线53完成相互间的通信;

[0148] 可选的,通信接口54可以为通信模块的接口,如GSM模块的接口。

[0149] 可选的,本发明实施例还提供了一种可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现如上述任一所述的关键词权重获取方法包含的各个步骤。

[0150] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。对于装置或系统类实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0151] 还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0152] 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块,或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0153] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

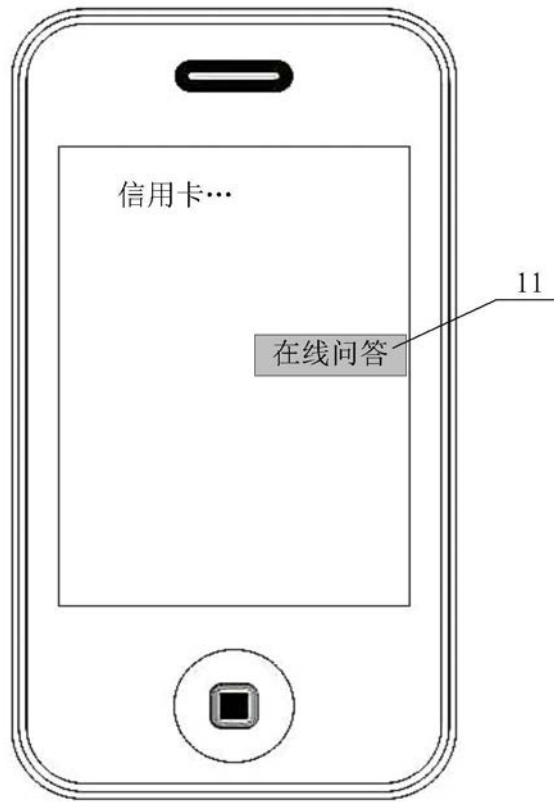


图1a

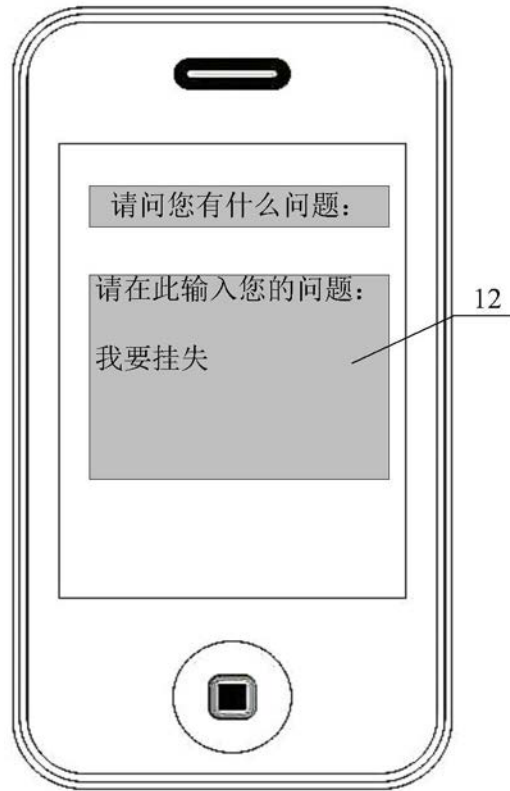


图1b

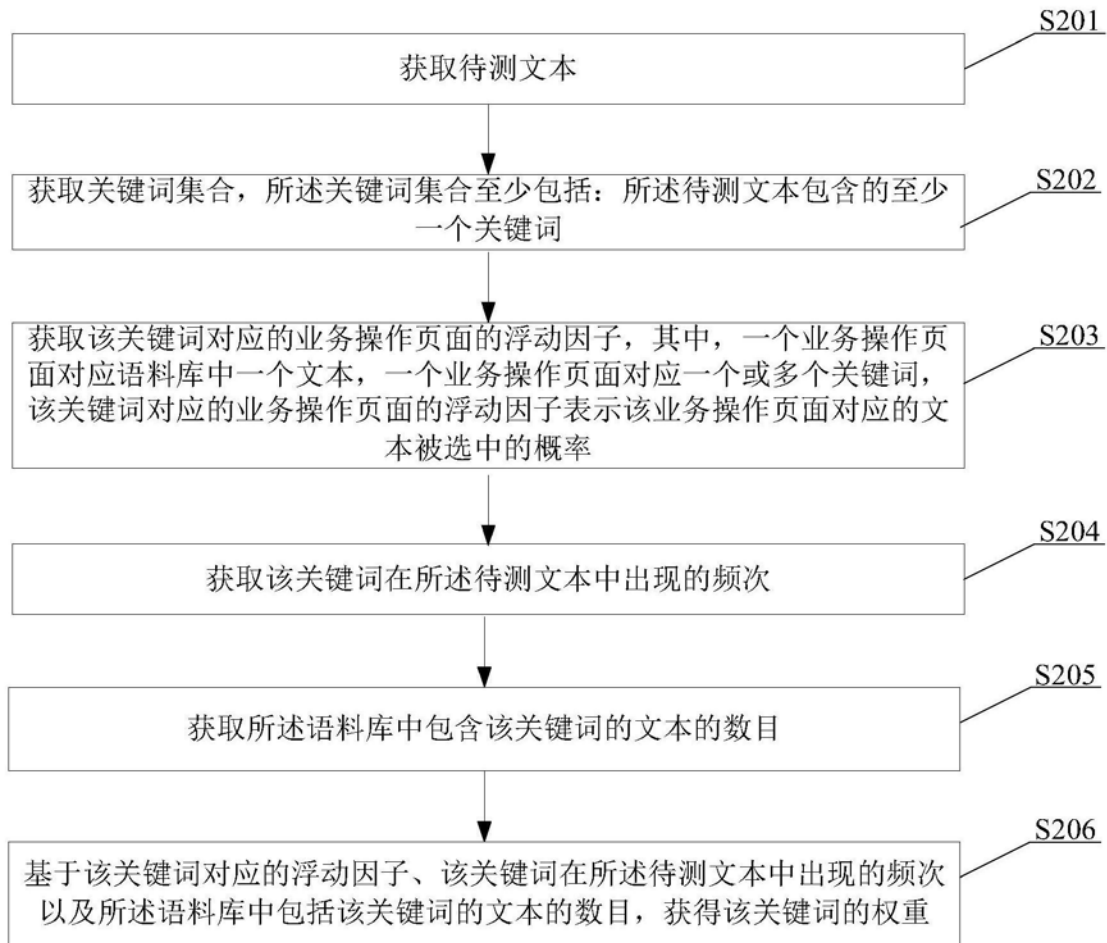


图2

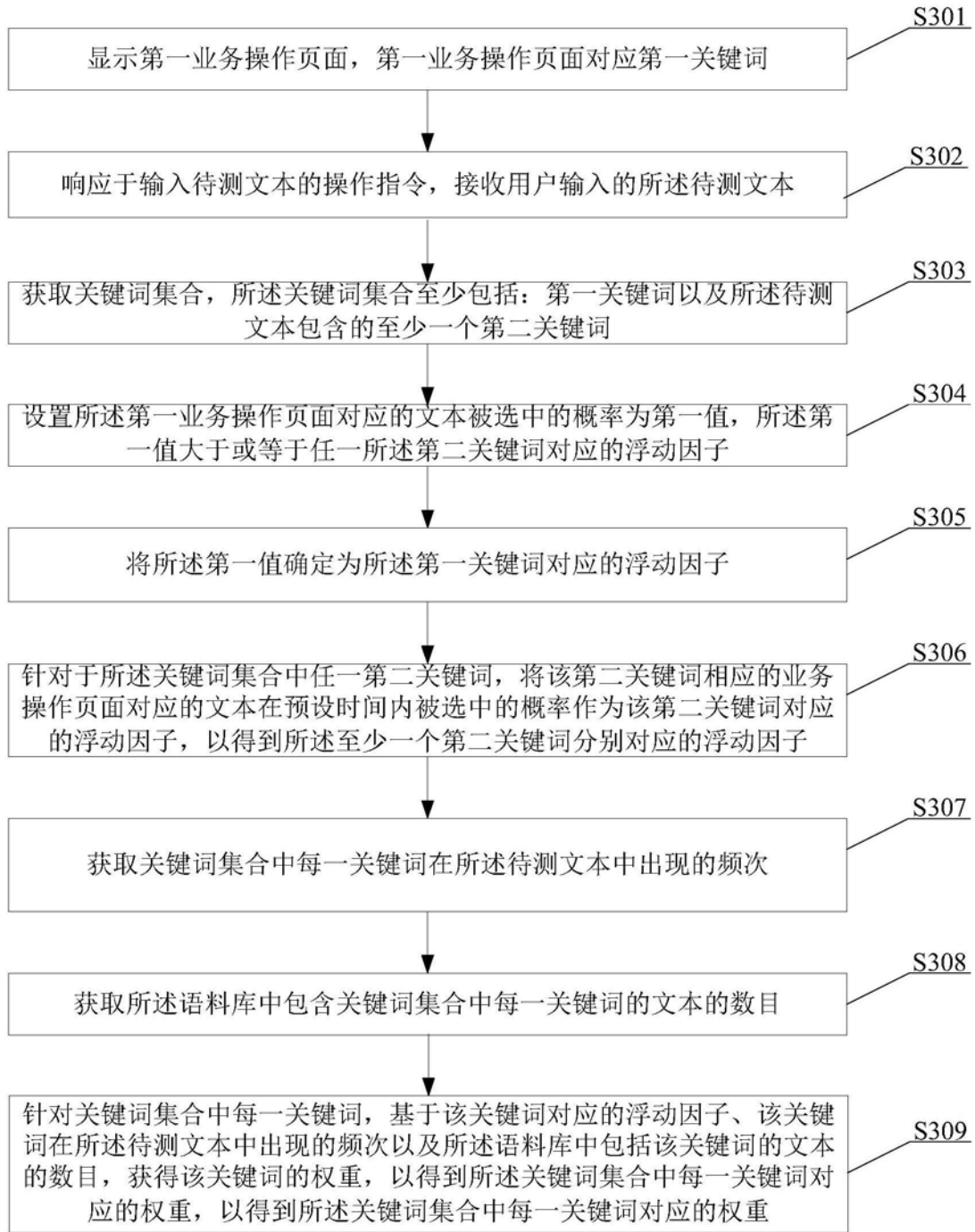


图3



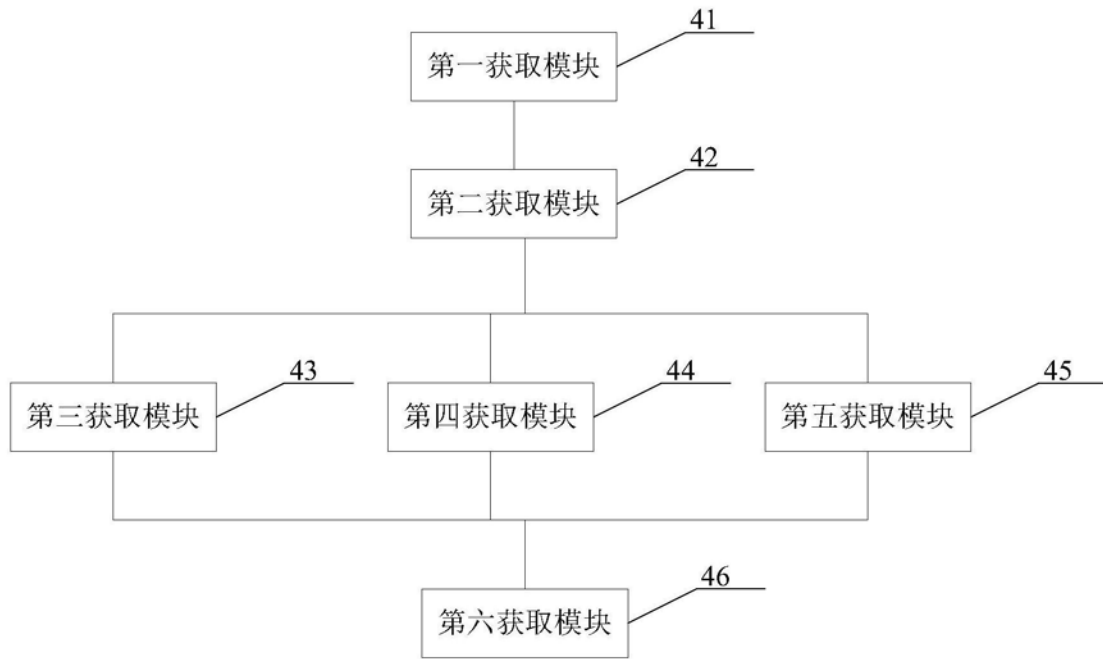


图4

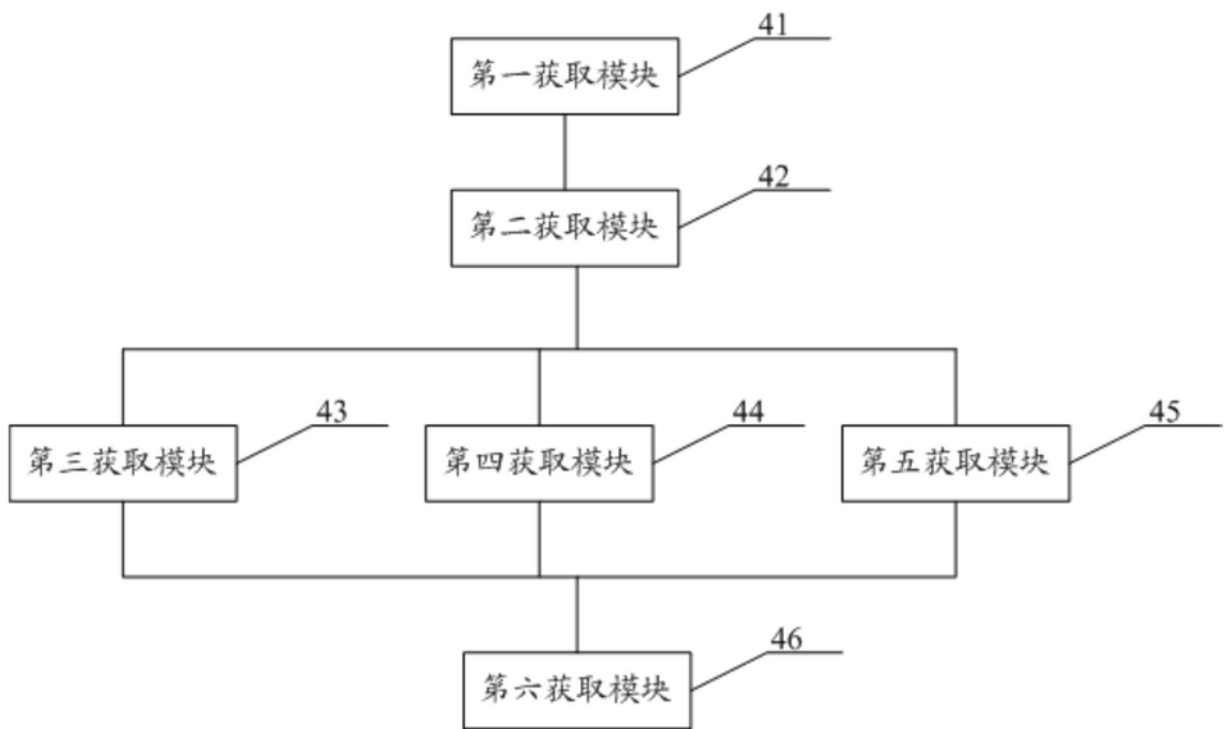


图5