(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113515423 B
(45)授权公告日 2023.05.23

(21)申请号 202110430544.1
(22)申请日 2021.04.21
(65)同一申请的已公布的文献号
申请公布号 CN 113515423 A
(43)申请公布日 2021.10.19
(73)专利权人 香港理工大学深圳研究院
地址 518057 广东省深圳市南山区粤海街道高新技术产业园南区粤兴一道18号香港理工大学产学研大楼205室
(72)发明人 马超 李俊彤 曹建农
(74)专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268
专利代理师 刘文求 朱阳波
(51)Int.Cl.
G06F 11/30(2006.01)
G06F 21/55(2013.01)
G06F 40/284(2020.01)
G06F 16/31(2019.01)

(56)对比文件
CN 103530365 A,2014.01.22
CN 104820588 A,2015.08.05
CN 106603327 A,2017.04.26
CN 109492150 A,2019.03.19
CN 111327762 A,2020.06.23
TW 201830326 A,2018.08.16
TW 202016693 A,2020.05.01
李子清.基于函数调用图的android恶意代码检测方法研究.《计算机测量与控制》.2017,全文.

Shu, JL等.Burn After Reading: Expunging Execution Footprints of Android Apps.《Web of Science》.2018,全文.

Yang, L等.A Hybrid Method for achieving High Accuracy and Eff iciency in Object Tracking using Passive RFID.《Web of Science》.2012,全文.

Li Liu等.Design and implementation of Android Phone Based Group Communication and Navigation System.《IEEE》.2012,全文.
审查员 文燕

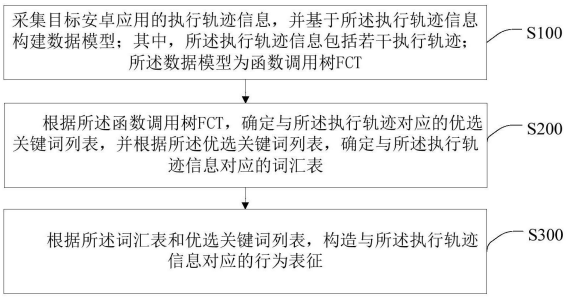
权利要求书2页 说明书11页 附图4页

(54)发明名称

基于执行轨迹信息的安卓应用行为表征构造方法

(57)摘要

本发明公开了基于执行轨迹信息的安卓应用行为表征构造方法,方法包括:采集目标安卓应用的执行轨迹信息,并基于执行轨迹信息构建数据模型;其中,执行轨迹信息包括若干执行轨迹;数据模型为函数调用树FCT;根据函数调用树FCT,确定与执行轨迹对应的优选关键词列表,并根据优选关键词列表,确定与执行轨迹信息对应的词汇表;根据词汇表和优选关键词列表,构造与执行轨迹信息对应的行为表征。本发明申请根据目标安卓应用的执行轨迹信息进行建模,根据建模后的数据模型来提取关键词,根据关键词来构造语义一致性的安卓应用行为表征,当将安卓应用行为表征应用于安卓应用功能识别时,可以显著提高安卓应用功能识别准确率。



CN 113515423 B

1.基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述方法包括:

采集目标安卓应用的执行轨迹信息,并基于所述执行轨迹信息构建数据模型;其中,所述执行轨迹信息包括若干执行轨迹;所述数据模型为函数调用树FCT;

根据所述函数调用树FCT,确定与所述执行轨迹对应的优选关键词列表,并根据所述优选关键词列表,确定与所述执行轨迹信息对应的词汇表;

根据所述词汇表和优选关键词列表,构造与所述执行轨迹信息对应的行为表征;

所述根据所述函数调用树FCT,确定与所述执行轨迹对应的优选关键词列表,并根据所述优选关键词列表,确定与所述执行轨迹信息对应的词汇表包括:

获取初始化词汇表;

获取初始化关键词列表;

初始化并获取轨迹索引值;

针对每一个与所述执行轨迹对应的函数调用树FCT,根据所述函数调用树FCT和初始化关键词列表,确定与所述执行轨迹对应的关键词列表和优选关键词列表,并更新所述轨迹索引值;

当所述轨迹索引值小于预设的轨迹索引阈值时,继续执行获取初始化关键词列表;初始化并获取轨迹索引值;针对每一个与所述执行轨迹对应的函数调用树FCT,根据所述函数调用树FCT和初始化关键词列表,确定与所述执行轨迹对应的关键词列表和优选关键词列表,并更新所述轨迹索引值的步骤;

当所述轨迹索引值大于或等于预设的轨迹索引阈值时,则输出与所述执行轨迹对应的优选关键词列表,并停止更新所述轨迹索引值;

根据所述优选关键词列表和所述初始化词汇表,确定与所述执行轨迹信息对应的词汇表。

2.根据权利要求1所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述基于所述执行轨迹信息构建数据模型包括:

获取初始化数据模型;

根据所述执行轨迹信息和初始化数据模型,循环构建数据模型。

3.根据权利要求2所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,根据所述执行轨迹信息和初始化数据模型,循环构建数据模型包括:

针对若干所述执行轨迹中的每一个所述执行轨迹,获取所述执行轨迹中的若干调用函数;

初始化并获取函数索引值;

针对若干调用函数中的每一个调用函数,根据所述调用函数和初始化数据模型,确定与所述调用函数对应的节点编号和父节点属性值,并更新所述函数索引值;

当所述函数索引值小于预设的函数索引阈值时,继续执行根据所述执行轨迹,确定与所述调用函数对应的节点编号和父节点属性值,并更新所述函数索引值的步骤;

当所述函数索引值大于或等于预设的函数索引阈值时,输出与所述执行轨迹对应的函数调用树FCT,并停止更新所述函数索引值。

4.根据权利要求3所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述更新所述函数索引值具体为:

将所述函数索引值加上第一预设值,得到中间函数索引值;

将所述中间函数索引值作为更新后的所述函数索引值。

5.根据权利要求1所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述更新所述轨迹索引值具体为:

将所述轨迹索引值加上第二预设值,得到中间轨迹索引值;

将所述中间轨迹索引值作为更新后的所述轨迹索引值。

6.根据权利要求5所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述根据所述词汇表和优选关键词列表,构造与所述执行轨迹信息对应的行为表征包括:

初始化并获取与所述执行轨迹对应的关键词索引值;

获取与所述执行轨迹对应的初始行为表征;

获取与所述词汇表对应的N维向量表征样本;其中,N为大于等于2的自然数;

根据所述词汇表和所述N维向量表征样本,对预设的原始模型进行训练,得到word2vec模型;

针对所述执行轨迹信息中的每一个所述执行轨迹,根据所述关键词索引值、所述优选关键词列表和所述word2vec模型,得到与所述执行轨迹对应的N维向量表征;

根据所述初始行为表征和所述N维向量表征,构造与所述执行轨迹信息对应的行为表征。

7.根据权利要求6所述的基于执行轨迹信息的安卓应用行为表征构造方法,其特征在于,所述针对所述执行轨迹信息中的每一个所述执行轨迹,根据所述关键词索引值、所述优选关键词列表和所述word2vec模型,得到与所述执行轨迹对应的N维向量表征包括:

针对所述执行轨迹信息中的每一个所述执行轨迹,将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型,得到与所述关键词索引值对应的中间N维向量表征;

当所述关键词索引值小于预设的关键词索引阈值时,继续执行针对所述执行轨迹信息中的每一个所述执行轨迹,将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型,得到与所述关键词索引值对应的中间N维向量表征的步骤,并更新所述关键词索引值;其中,所述更新所述关键词索引值为将所述关键词索引值加上第三预设值的结果作为更新后的所述关键词索引值;

当所述关键词索引值大于或等于预设的关键词索引阈值时,将所有所述中间N维向量表征进行水平拼接,得到与所述执行轨迹对应的N维向量表征。

8.一种智能终端,其特征在于,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如权利要求1-7中任意一项所述的方法。

9.一种非临时性计算机可读存储介质,其特征在于,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如权利要求1-7中任意一项所述的方法。

基于执行轨迹信息的安卓应用行为表征构造方法

技术领域

[0001] 本发明涉及信息技术领域，尤其涉及的是基于执行轨迹信息的安卓应用行为表征构造方法。

背景技术

[0002] 随着智能移动终端尤其是基于Android操作系统的智能手机在全球范围内占据最大市场份额，准确识别Android应用的功能对软件测试维护、恶意软件检测、用户隐私保护等任务都具有重要作用，由于Android运行环境的动态性和开发框架的差异性，Android应用的执行轨迹非常复杂，在记录规模和行为模式方面都体现出巨大的区别，基于执行轨迹的安卓应用行为表征构造方法是影响Android应用功能识别效果的重要因素，但是现有技术的Android应用行为表征构造方法都是基于程序静态分析，无法准确捕获Android应用动态行为，使得将安卓应用行为表征应用于安卓应用功能识别时准确率低。

[0003] 因此，现有技术还有待改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供基于执行轨迹信息的安卓应用行为表征构造方法，旨在解决现有技术中Android应用行为表征构造方法都是基于程序静态分析，无法准确捕获Android应用动态行为，使得将安卓应用行为表征应用于安卓应用功能识别时准确率低的问题。

[0005] 本发明解决问题所采用的技术方案如下：

[0006] 第一方面，本发明实施例提供基于执行轨迹信息的安卓应用行为表征构造方法，其中，所述方法包括：

[0007] 采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；

[0008] 根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；

[0009] 根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征。

[0010] 在一种实现方式中，其中，所述基于所述执行轨迹信息构建数据模型包括：

[0011] 获取初始化数据模型；

[0012] 根据所述执行轨迹信息和初始化数据模型，循环构建数据模型。

[0013] 在一种实现方式中，其中，所述根据所述执行轨迹信息和初始化数据模型，循环构建数据模型包括：

[0014] 针对若干所述执行轨迹中的每一个所述执行轨迹，获取所述执行轨迹中的若干调用函数；

[0015] 初始化并获取函数索引值；

[0016] 针对若干调用函数中的每一个调用函数，根据所述调用函数和初始化数据模型，

确定与所述调用函数对应的节点编号和父节点属性值，并更新所述函数索引值；

[0017] 当所述函数索引值小于预设的函数索引阈值时，继续执行根据所述执行轨迹，确定与所述调用函数对应的节点编号和父节点属性值，并更新所述函数索引值的步骤；

[0018] 当所述函数索引值大于或等于预设的函数索引阈值时，输出与所述执行轨迹对应的函数调用树FCT，并停止更新所述函数索引值。

[0019] 在一种实现方式中，其中，所述更新所述函数索引值具体为：

[0020] 将所述函数索引值加上第一预设值，得到中间函数索引值；

[0021] 将所述中间函数索引值作为更新后的所述函数索引值。

[0022] 在一种实现方式中，其中，所述根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表包括：

[0023] 获取初始化词汇表；

[0024] 获取初始化关键词列表；

[0025] 初始化并获取轨迹索引值；

[0026] 针对每一个与所述执行轨迹对应的函数调用树FCT，根据所述函数调用树FCT和初始化关键词列表，确定与所述执行轨迹对应的关键词列表和优选关键词列表，并更新所述轨迹索引值；

[0027] 当所述轨迹索引值小于预设的轨迹索引阈值时，继续执行获取初始化关键词列表；初始化并获取轨迹索引值；针对每一个与所述执行轨迹对应的函数调用树FCT，根据所述函数调用树FCT和初始化关键词列表，确定与所述执行轨迹对应的关键词列表和优选关键词列表，并更新所述轨迹索引值的步骤；

[0028] 当所述轨迹索引值大于或等于预设的轨迹索引阈值时，则输出与所述执行轨迹对应的优选关键词列表，并停止更新所述轨迹索引值；

[0029] 根据所述优选关键词列表和所述初始化词汇表，确定与所述执行轨迹信息对应的词汇表。

[0030] 在一种实现方式中，其中，所述更新所述轨迹索引值具体为：

[0031] 将所述轨迹索引值加上第二预设值，得到中间轨迹索引值；

[0032] 将所述中间轨迹索引值作为更新后的所述轨迹索引值。

[0033] 在一种实现方式中，其中，所述根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征包括：

[0034] 初始化并获取与所述执行轨迹对应的关键词索引值；

[0035] 获取与所述执行轨迹对应的初始行为表征；

[0036] 获取与所述词汇表对应的N维向量表征样本；其中，N为大于等于2的自然数；

[0037] 根据所述词汇表和所述N维向量表征样本，对预设的原始模型进行训练，得到word2vec模型；

[0038] 针对所述执行轨迹信息中的每一个所述执行轨迹，根据所述关键词索引值、所述优选关键词列表和所述word2vec模型，得到与所述执行轨迹对应的N维向量表征；

[0039] 根据所述初始行为表征和所述N维向量表征，构造与所述执行轨迹信息对应的行为表征。

[0040] 在一种实现方式中，其中，所述针对所述执行轨迹信息中的每一个所述执行轨迹，根据所述关键词索引值、所述优选关键词列表和所述word2vec模型，得到与所述执行轨迹对应的N维向量表征包括：

[0041] 针对所述执行轨迹信息中的每一个所述执行轨迹，将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型，得到与所述关键词索引值对应的中间N维向量表征；

[0042] 当所述关键词索引值小于预设的关键词索引阈值时，继续执行针对所述执行轨迹信息中的每一个所述执行轨迹，将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型，得到与所述关键词索引值对应的中间N维向量表征的步骤，并更新所述关键词索引值；其中，所述更新所述关键词索引值为将所述关键词索引值加上第三预设值的结果作为更新后的所述关键词索引值；

[0043] 当所述关键词索引值大于或等于预设的关键词索引阈值时，将所有所述中间N维向量表征进行水平拼接，得到与所述执行轨迹对应的N维向量表征。

[0044] 第二方面，本发明实施例还提供基于执行轨迹信息的安卓应用行为表征构造装置，其中，所述装置包括：

[0045] 数据模型构建单元，用于采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；

[0046] 优选关键词列表和词汇表的获取单元，用于根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；

[0047] 行为表征构造单元，用于根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征。

[0048] 第三方面，本发明实施例还提供一种智能终端，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于执行如上述任意一项所述的基于执行轨迹信息的安卓应用行为表征构造方法。

[0049] 第四方面，本发明实施例还提供一种非临时性计算机可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如上述中任意一项所述的基于执行轨迹信息的安卓应用行为表征构造方法。

[0050] 本发明的有益效果：本发明实施例首先采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；然后根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；最后根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征；可见，本发明实施例中根据目标安卓应用的执行轨迹信息进行建模，根据建模后的数据模型来提取关键词，根据关键词来构造语义一致性的安卓应用行为表征，当将安卓应用行为表征应用于安卓应用功能识别时，可以显著提高安卓应用功能识别准确率。

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中记载的一些实施例，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。

[0052] 图1为本发明实施例提供的基于执行轨迹信息的安卓应用行为表征构造方法流程示意图。

[0053] 图2为本发明实施例提供的Android应用执行轨迹信息采集及建模流程图。

[0054] 图3为本发明实施例提供的Android应用执行轨迹关键词提取及词汇表构建流程图。

[0055] 图4为本发明实施例提供的Android应用执行轨迹行为表征构造流程图。

[0056] 图5为本发明实施例提供的基于执行轨迹信息的安卓应用行为表征构造装置的原理框图。

[0057] 图6为本发明实施例提供的智能终端的内部结构原理框图。

具体实施方式

[0058] 本发明公开了基于执行轨迹信息的安卓应用行为表征构造方法、智能终端、存储介质，为使本发明的目的、技术方案及效果更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定本发明。

[0059] 本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解，当我们称元件被“连接”或“耦接”到另一元件时，它可以直接连接或耦接到其他元件，或者也可以存在中间元件。此外，这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

[0060] 本技术领域技术人员可以理解，除非另外定义，这里使用的所有术语(包括技术术语和科学术语)，具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是，诸如通用字典中定义的那些术语，应该被理解为具有与现有技术的上下文中的意义一致的意义，并且除非像这里一样被特定定义，否则不会用理想化或过于正式的含义来解释。

[0061] 由于现有技术中，Android应用行为表征构造方法都是基于程序静态分析，无法准确捕获Android应用动态行为，使得将安卓应用行为表征应用于安卓应用功能识别时准确率低的问题。

[0062] 为了解决现有技术的问题，本实施例提供了基于执行轨迹信息的安卓应用行为表征构造方法，本发明根据目标安卓应用的执行轨迹信息进行建模，根据建模后的数据模型来提取关键词，根据关键词来构造语义一致性的安卓应用行为表征，当将安卓应用行为表征应用于安卓应用功能识别时，可以显著提高安卓应用功能识别准确率。具体实施时，首先

采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；然后根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；最后根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征。

[0063] 示例性方法

[0064] 本实施例提供基于执行轨迹信息的安卓应用行为表征构造，该方法可以应用于信息技术智能终端。具体如图1所示，所述方法包括：

[0065] 步骤S100、采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；

[0066] 具体地，本发明采用程序插装技术采集目标安卓应用的执行轨迹信息，其中，所述执行轨迹信息包括若干执行轨迹。例如：采用程序插装技术采集目标安卓应用的执行轨迹$x_i=\{mc_j\}$ $(2\leqslant i)$，所有的执行轨迹x_i组成执行轨迹信息。其中mc_j为执行轨迹x_i中的按照时间顺序在第j个调用的函数$(1\leqslant j\leqslant|x_i|)$；其中，$|x_i|$表示执行轨迹$x_i$中函数的个数。程序插装技术为借助往被测程序中插入操作(称为“探针”)，以便获取程序的控制流和数据流信息，从而实现测试目的的方法。在软件动态测试中，程序插装是一种基本的测试手段，应用广泛，是覆盖率测试、软件故障注入和动态性能分析的基础技术。然后根据所述执行轨迹信息构建统一的数据模型，在本实施例中，数据模型为函数调用树FCT，为后续构造具有语义一致性的Android应用行为表征做准备。

[0067] 在本发明实施例的一种实现方式中，所述基于所述执行轨迹信息构建数据模型包括如下步骤：获取初始化数据模型；根据所述执行轨迹信息和初始化数据模型，循环构建数据模型。

[0068] 具体地，先获取初始化数据模型，实际中，针对执行轨迹信息中的每一个执行轨迹，都要初始化执行轨迹x_i的函数调用树$FCT(x_i)$，由于每个函数调用树$FCT(x_i)$中根节点对应第一个调用函数mc_1，由于调用函数mc_1没有主调函数，故将该根节点的父节点的节点编号设置为0，将该父节点的父节点属性值设置为NULL，通过上述设置就可以得到初始化数据模型，系统就可以获取到该初始化数据模型，然后根据所述执行轨迹信息和初始化数据模型，循环构建数据模型。相应的，所述根据所述执行轨迹信息和初始化数据模型，循环构建数据模型包括如下步骤：针对若干所述执行轨迹中的每一个所述执行轨迹，获取所述执行轨迹中的若干调用函数；初始化并获取函数索引值；针对若干调用函数中的每一个调用函数，根据所述调用函数和初始化数据模型，确定与所述调用函数对应的节点编号和父节点属性值，并更新所述函数索引值；当所述函数索引值小于预设的函数索引阈值时，继续执行根据所述执行轨迹，确定与所述调用函数对应的节点编号和父节点属性值，并更新所述函数索引值的步骤；当所述函数索引值大于或等于预设的函数索引阈值时，输出与所述执行轨迹对应的函数调用树FCT，并停止更新所述函数索引值。

[0069] 在本实施例中，如图2所示，针对若干所述执行轨迹中的每一个所述执行轨迹，获取所述执行轨迹中的若干调用函数，并初始化函数索引值，实际中，每一个执行轨迹都对应一个函数索引值。初始化数据模型后，需要确定从每个执行轨迹的第一调用函数mc_1到每个执行轨迹的第j个调用函数mc_j的节点编号和父节点属性值，当执行轨迹中的调用函数mc_j的

主调函数为mc_s，且$mc_s \in x_i$时，第j个调用函数mc_j的节点编号为j，父节点属性值为S，并更新所述函数索引值；相应的，所述更新所述函数索引值具体为：将所述函数索引值加上第一预设值，得到中间函数索引值；将所述中间函数索引值作为更新后的所述函数索引值。在本实施例中，第一预设值取值为1。当所述函数索引值小于预设的函数索引阈值时，如$j<|x_i|+1$（在本实施例中，预设的函数索引阈值为执行轨迹中调用函数的个数加1后的值），更新所述函数索引值为j＝j+1，继续执行根据所述执行轨迹，确定与所述调用函数对应的节点编号和父节点属性值的步骤；当所述函数索引值大于或等于预设的函数索引阈值时，输出与所述执行轨迹对应的函数调用树FCT，并停止更新所述函数索引值。

[0070] 现举例说明基于所述执行轨迹信息构建数据模型的过程：

[0071] (1)利用程序插装技术获得目标Android应用的执行轨迹信息集合$X=\{x_1,x_2\}$，其中执行轨迹x_1＝{‘ViewMemo’，‘CreateString’}中的调用函数‘ViewMemo’为调用函数‘CreateString’的主调函数，执行轨迹x_2＝{‘EditMemo’，‘ViewItem’}中的调用函数‘EditMemo’为函数‘ViewItem’的主调函数；

[0072] (2)初始化赋值执行轨迹x_1的函数调用树$FCT(x_1)$的根节点对应mc_1，由于mc_1的主调函数为空(因为函数‘ViewMemo’无主调函数)，则将mc_1对应的根节点的父节点的编号设置为0，函数调用树$FCT(x_1)$根节点的父节点的父节点属性赋值为Null；

[0073] (3)获取执行轨迹x_1的第一个调用函数‘ViewMemo’，调用函数‘ViewMemo’在$FCT(x_1)$的中为根节点，该根节点的节点编号设置为1，该根节点的父节点属性值为0；

[0074] (4)更新函数索引$j=1+1=2<|x_1|+1$，其中，$|x_1|$的值为2，因此执行轨迹x_1的第二个调用函数‘CreateString’；由于调用函数‘CreateString’的主调函数为‘ViewMemo’$\in x_1$，因此将函数调用树$FCT(x_1)$的当前节点$FCT(x_1)_2$节点编号设置为2，该当前节点$FCT(x_1)_2$的父节点属性设置为1(因为函数‘CreateString’在执行轨迹x_1中的主调函数‘ViewMemo’的序号为1)；

[0075] (5)更新函数索引$j=2+1=3=|x_1|+1$，其中，$|x_1|$的值为2，因此执行环节(6)；

[0076] (6)输出执行轨迹x_1的函数调用树$FCT(x_1)$。

[0077] 对于执行轨迹x_2重复执行上述流程可获得其对应的函数调用树$FCT(x_2)$。

[0078] 得到数据模型也即函数调用树FCT后，就可以执行如如1所示的如下步骤：S200、根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；相应的，所述根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表包括如下步骤：

[0079] S201、获取初始化词汇表；

[0080] S202、获取初始化关键词列表；

[0081] S203、初始化并获取轨迹索引值；

[0082] S204、针对每一个与所述执行轨迹对应的函数调用树FCT，根据所述函数调用树FCT和初始化关键词列表，确定与所述执行轨迹对应的关键词列表和优选关键词列表，并更新所述轨迹索引值；

[0083] S205、当所述轨迹索引值小于预设的轨迹索引阈值时，继续执行获取初始化关键词列表；初始化并获取轨迹索引值；针对每一个与所述执行轨迹对应的函数调用树FCT，根

据所述函数调用树FCT和初始化关键词列表，确定与所述执行轨迹对应的关键词列表和优选关键词列表，并更新所述轨迹索引值的步骤；

[0084] S206、当所述轨迹索引值大于或等于预设的轨迹索引阈值时，则输出与所述执行轨迹对应的优选关键词列表，并停止更新所述轨迹索引值；

[0085] S207、根据所述优选关键词列表和所述初始化词汇表，确定与所述执行轨迹信息对应的词汇表。

[0086] 具体地，如图3所示，先获取初始化词汇表和初始化关键词列表；例如，先初始化执行轨迹集合$X=\{x_i\}$对应的词汇表$V=\{\}$；然后初始化执行轨迹x_i对应的关键词列表$KW(x_i)=\{\}$；然后初始化并获取轨迹索引值，其中，每一个执行轨迹对应一个轨迹索引值；将轨迹索引值初始化为0。接着获取初始化词汇表$V=\{\}$和初始化关键词列表$KW(x_i)=\{\}$。针对每一个与所述执行轨迹对应的函数调用树FCT，根据所述函数调用树FCT，执行轨迹x_i函数调用树节点$FCT(x_i)_j$的关键词列表$KW(x_i)_{mcj}=Camel_Case(mc_j)$，其中，$Camel_Case(mc_j)$为利用驼峰规则从函数$mc_j$的函数名中以大写字母为开始标志，以下一个大写字母的前一个小写字母为结束标志；将函数mc_j的函数名进行切分，得到候选关键词列表集合$KW(x_i)_{mcj}$；然后以宽度优先遍历的方式遍历执行轨迹x_i的函数调用树FCT的所有节点，并获得所有节点的关键词列表，将所有的关键词列表进行合并，得到关键词列表$KW(x_i)=\cup_{1\leqslant j\leqslant|xi|}KW(x_i)_{mcj}$；对执行轨迹$x_i$的关键词列表$KW(x_i)$中每个关键词，计算该关键词的tf和idf值(词频TF表示词条(关键字)在文本中出现的频率，逆向文件频率(IDF)由总文件数目除以包含该词语的文件的数目得到商，再将得到的商取对数得到)，然后将该关键词的tf与idf值进行相乘得到乘积值，再将乘积值进行从高到低的排序，取排序最高的k个关键词构成中间优选关键词列表$KW(x_i,k)$；将所述中间优选关键词列表$KW(x_i,k)$进行合并$\cup_{1\leqslant i\leqslant|xi|}KW(x_i,k)$得到优选关键词列表KW(k)；此时，更新所述轨迹索引值，相应的所述更新所述轨迹索引值具体为：将所述轨迹索引值加上第二预设值，得到中间轨迹索引值；将所述中间轨迹索引值作为更新后的所述轨迹索引值。在本实施例中，第二预设值为1，更新所述轨迹索引值为$i=i+1$。当所述轨迹索引值小于预设的轨迹索引阈值，如：$i<|X|+1$(在本实施例中，轨迹索引阈值为集合X中元素的个数加1后的和)时，继续执行获取初始化关键词列表；初始化并获取轨迹索引值；针对每一个与所述执行轨迹对应的函数调用树FCT，根据所述函数调用树FCT和初始化关键词列表，确定与所述执行轨迹对应的关键词列表和优选关键词列表，并更新所述轨迹索引值的步骤；当所述轨迹索引值大于或等于预设的轨迹索引阈值时，则输出与所述执行轨迹对应的优选关键词列表KW(k)，并停止更新所述轨迹索引值；根据所述优选关键词列表和所述初始化词汇表，确定与所述执行轨迹信息对应的词汇表，也即$V=V\cup KW(k)$。

[0087] 现举例说明优选关键词列表和词汇表的确定过程，如下步骤：

[0088] (1)初始化执行轨迹集合$X=\{x_1,x_2\}$对应的词汇表$V=\{\}$；

[0089] (2)初始化执行轨迹x_1对应的关键词列表$KW(x_1)=\{\}$；

[0090] (3)构造执行轨迹x_1函数调用树节点$FCT(x_1)_1$的关键词列表$KW(x_1)_{'ViewMemo'}=\{'View','Memo'\}$；

[0091] (4)以宽度优先遍历的方式遍历执行轨迹x_1函数调用树的所有节点，并获得其关键词列表的并集$KW(x_i)=\{'View','Memo'\}\cup\{'Create','String'\}$；

[0092] (5)利用tf-idf模型对执行轨迹x_1的关键词列表$KW(x_1)$中所有的关键词进行排序

并返回排序最高的2个(假设k＝2)中间优选关键词列表KW(x_1,2)＝{'Memo'，'Create'}；

[0093] (6)更新执行轨迹索引i＝1+1＝2<|X|+1,因此返回环节(2)；

[0094] (7)重复执行环节(2)-(6),从而获得执行轨迹x_2中排序最高的2个中间优选关键词列表KW(x_2,2)＝{'Memo'，'View'},并更新执行对应的优选关键词列表KW(2)＝KW(x_1,2)({'Memo'，'Create'})∪KW(x_1,2)({'Memo'，'View'})＝{'Memo'，'View'，'Create'}；

[0095] (8)更新执行轨迹索引i＝2+1＝3＝|X|+1(|X|的值为2),因此执行环节(9)；

[0096] (9)输出执行轨迹集合X的词汇表V＝{}∪{'Memo'，'View'，'Create'}＝{'Memo'，'View'，'Create'}。

[0097] 得到所述词汇表和优选关键词列表后,就可以执行如图1所示的如下步骤:S300、根据所述词汇表和优选关键词列表,构造与所述执行轨迹信息对应的行为表征。相应的,所述根据所述词汇表和优选关键词列表,构造与所述执行轨迹信息对应的行为表征包括如下步骤：

[0098] S301、初始化并获取与所述执行轨迹对应的关键词索引值；

[0099] S302、获取与所述执行轨迹对应的初始行为表征；

[0100] S303、获取与所述词汇表对应的N维向量表征样本;其中,N为大于等于2的自然数；

[0101] S304、根据所述词汇表和所述N维向量表征样本,对预设的原始模型进行训练,得到word2vec模型；

[0102] S305、针对所述执行轨迹信息中的每一个所述执行轨迹,根据所述关键词索引值、所述优选关键词列表和所述word2vec模型,得到与所述执行轨迹对应的N维向量表征；

[0103] S306、根据所述初始行为表征和所述N维向量表征,构造与所述执行轨迹信息对应的行为表征。

[0104] 具体地,如图4所示,初始化并获取与所述执行轨迹对应的关键词索引值;在本实施例中,关键词索引值p的初始值为0;然后获取与所述执行轨迹对应的初始行为表征SR(x_1)＝<>;接着获取与所述词汇表对应的N维向量表征样本word2vec(V,d);其中,N为大于等于2的自然数;将所述词汇表输入至预设的原始模型进行训练,输出原始模型输出数据,根据原始模型输出数据和N维向量表征样本得到损失函数,根据所述损失函数调整原始模型的参数,当满足预设的条件时,则停止训练,得到word2vec模型。再针对所述执行轨迹信息中的每一个所述执行轨迹,根据所述关键词索引值、所述优选关键词列表和所述word2vec模型,得到与所述执行轨迹对应的N维向量表征,相应的,所述针对所述执行轨迹信息中的每一个所述执行轨迹,根据所述关键词索引值、所述优选关键词列表和所述word2vec模型,得到与所述执行轨迹对应的N维向量表征包括如下步骤:针对所述执行轨迹信息中的每一个所述执行轨迹,将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型,得到与所述关键词索引值对应的中间N维向量表征;当所述关键词索引值小于预设的关键词索引阈值时,继续执行针对所述执行轨迹信息中的每一个所述执行轨迹,将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型,得到与所述关键词索引值对应的中间N维向量表征的步骤,并更新所述关键词索引值;其中,所述更新所述关键词索引值为将所述关键词索引值加上第三预设值的结果作为更新后的所述关键词索引值;当所述关键词索引值大于或等于预设的关键词索引阈值时,将所有所述中间N维向量表征进行水平拼接,得到与所述执行轨迹对应的N维向量表征。

[0105] 具体地，将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型，在本实施例中，执行轨迹x_i中索引为p的关键词$kw_{i,p}$的N维向量表征getVector($kw_{i,p}$,vector_model)，其中getVector($kw_{i,p}$,vector_model)方法为基于N维向量模型vector_model＝word2vec(V,N)输出关键词$kw_{i,p}$对应的N维向量表征。当所述关键词索引值小于预设的关键词索引阈值，如:p<|KW(x_i,k)|+1(在本实施例中，关键词索引阈值为第i个执行轨迹选取的K个关键词组成集合的个数+1的和，如:|KW(x_i,k)|+1)时，继续执行针对所述执行轨迹信息中的每一个所述执行轨迹，将所述优选关键词列表中与关键词索引值对应的关键词输入至所述word2vec模型，得到与所述关键词索引值对应的中间N维向量表征的步骤，并更新所述关键词索引值；其中，所述更新所述关键词索引值为将所述关键词索引值加上第三预设值的结果作为更新后的所述关键词索引值，如p＝p+1；当所述关键词索引值大于或等于预设的关键词索引阈值时，将所有所述中间N维向量表征进行水平拼接，得到与所述执行轨迹对应的N维向量表征。最后，根据所述初始行为表征和所述N维向量表征，构造与所述执行轨迹信息对应的行为表征，也即将初始行为表征和N维向量表征进行水平拼接，构造与所述执行轨迹信息对应的行为表征。

[0106] 现举例说明行为表征构造的如下步骤：

[0107] (1)基于执行轨迹集合X的词汇表V训练3维向量模型vector_model＝word2vec(V,3)；

[0108] (2)初始化执行轨迹x_1对应的行为表征SR(x_1)＝<>；

[0109] (3)获取执行轨迹x_1中关键词‘ViewMemo’的3维向量表征(1,1,0)；

[0110] (4)更新关键词索引p＝1+1＝2<|KW(x_1,2)|+1，因此返回环节(3)；

[0111] (5)重复执行环节(3)-(4)，从而获得执行轨迹x_1中索引为2的关键词‘CreateString’对应的3维向量表征(0,0,1)；

[0112] (6)将执行轨迹x_1中关键词‘ViewMemo’的3维向量表征(1,1,0)与x_1中关键词‘CreateString’对应的3维向量表征(0,0,1)进行水平拼接Concatenate((1,1,0),(0,0,1))＝(1,1,0,0,0,1)

[0113] (7)更新关键词索引p＝2+1＝3＝|KW(x_1,2)|+1，因此执行环节(8)；

[0114] (8)输出执行轨迹x_1的行为表征SR(x_1)＝Concatenate(<>,1,1,0,0,0,1)＝(1,1,0,0,0,1)。

[0115] 对于执行轨迹x_2重复执行上述流程可获得其对应的行为表征SR(x_2)。构造好执行轨迹x_1和x_2的向量化表征后即可将其作为输入应用于现有基于机器学习模型的Android功能识别解决方案。

[0116] 示例性设备

[0117] 如图5中所示，本发明实施例提供基于执行轨迹信息的安卓应用行为表征构造装置，该装置包括数据模型构建单元401，优选关键词列表获取单元402，行为表征构造单元403，其中：

[0118] 数据模型构建单元401，用于采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；

[0119] 优选关键词列表和词汇表的获取单元402，用于根据所述函数调用树FCT，确定与

所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；

[0120] 行为表征构造单元403，用于根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征。

[0121] 基于上述实施例，本发明还提供了一种智能终端，其原理框图可以如图6所示。该智能终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏、温度传感器。其中，该智能终端的处理器用于提供计算和控制能力。该智能终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该智能终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现基于执行轨迹信息的安卓应用行为表征构造方法。该智能终端的显示屏可以是液晶显示屏或者电子墨水显示屏，该智能终端的温度传感器是预先在智能终端内部设置，用于检测内部设备的运行温度。

[0122] 本领域技术人员可以理解，图6中的原理图，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的智能终端的限定，具体的智能终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0123] 在一个实施例中，提供了一种智能终端，包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令：

[0124] 采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT；

[0125] 根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表；

[0126] 根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征。

[0127] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0128] 综上所述，本发明公开了基于执行轨迹信息的安卓应用行为表征构造方法、智能终端、存储介质，所述方法包括：采集目标安卓应用的执行轨迹信息，并基于执行轨迹信息构建数据模型；其中，执行轨迹信息包括若干执行轨迹；数据模型为函数调用树FCT；根据函数调用树FCT，确定与执行轨迹信息对应的优选关键词列表，并根据优选关键词列表，确定与执行轨迹信息对应的词汇表；根据词汇表和优选关键词列表，构造与执行轨迹信息对应的行为表征。本发明申请根据目标安卓应用的执行轨迹信息进行建模，根据建模后的数据

模型来提取关键词，根据关键词来构造语义一致性的安卓应用行为表征，当将安卓应用行为表征应用于安卓应用功能识别时，可以显著提高安卓应用功能识别准确率。

[0129] 基于上述实施例，本发明公开了基于执行轨迹信息的安卓应用行为表征构造方法，应当理解的是，本发明的应用不限于上述的举例，对本领域普通技术人员来说，可以根据上述说明加以改进或变换，所有这些改进和变换都应属于本发明所附权利要求的保护范围。

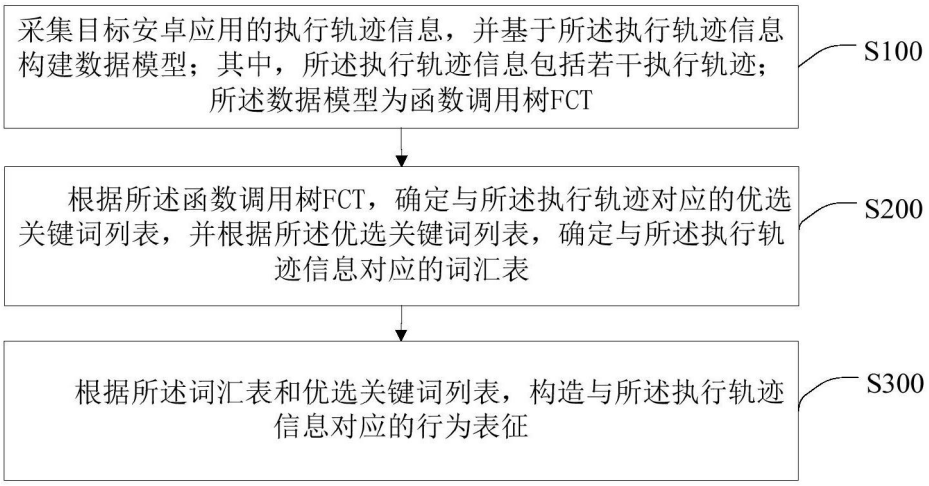

采集目标安卓应用的执行轨迹信息，并基于所述执行轨迹信息构建数据模型；其中，所述执行轨迹信息包括若干执行轨迹；所述数据模型为函数调用树FCT
S100
根据所述函数调用树FCT，确定与所述执行轨迹对应的优选关键词列表，并根据所述优选关键词列表，确定与所述执行轨迹信息对应的词汇表
S200
根据所述词汇表和优选关键词列表，构造与所述执行轨迹信息对应的行为表征
S300

图1

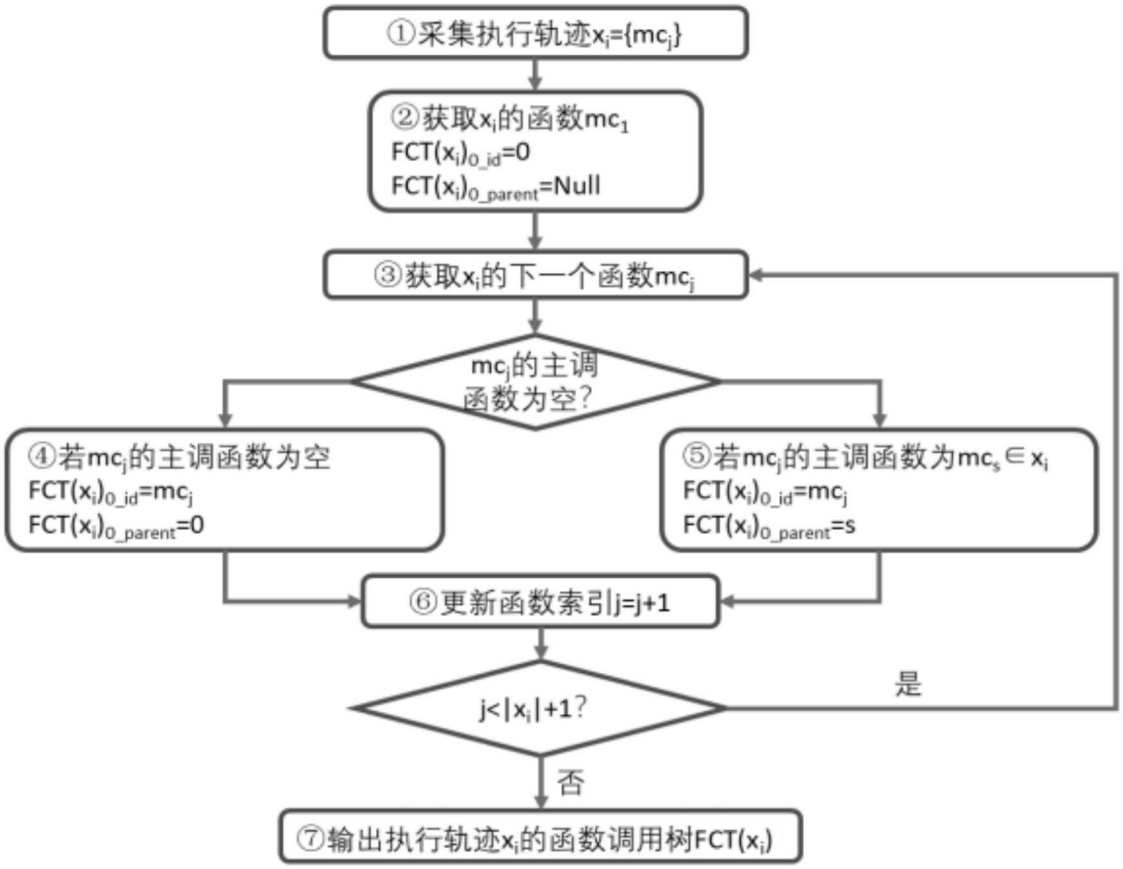

①采集执行轨迹$x_i=\{mc_j\}$
②获取x_i的函数mc_1
$FCT(x_i)_{0\_id}=0$
$FCT(x_i)_{0\_parent}=Null$
③获取x_i的下一个函数mc_j
mc_j的主调函数为空?
④若mc_j的主调函数为空
$FCT(x_i)_{0\_id}=mc_j$
$FCT(x_i)_{0\_parent}=0$
⑤若mc_j的主调函数为$mc_s \in x_i$
$FCT(x_i)_{0\_id}=mc_j$
$FCT(x_i)_{0\_parent}=s$
⑥更新函数索引j=j+1
$j<|x_i|+1$?
是
否
⑦输出执行轨迹x_i的函数调用树$FCT(x_i)$

图2

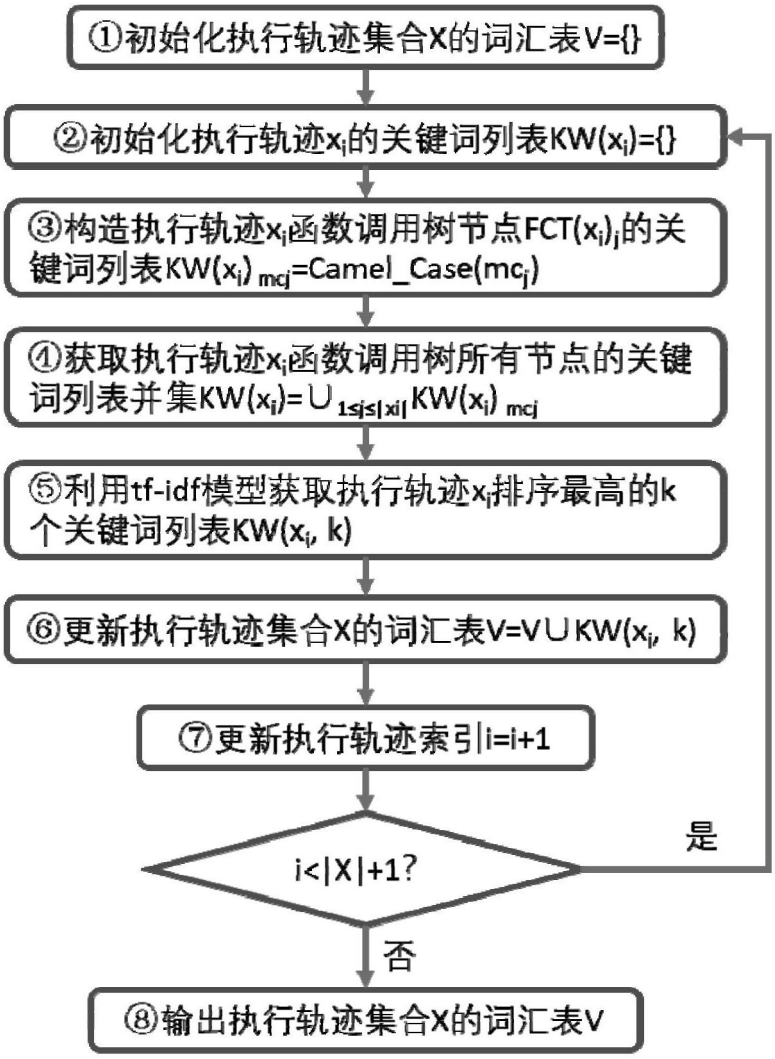

①初始化执行轨迹集合X的词汇表V={}
②初始化执行轨迹x_i的关键词列表$KW(x_i)$={}
③构造执行轨迹x_i函数调用树节点$FCT(x_i)_j$的关键词列表$KW(x_i)_{mcj}$=Camel_Case(mc_j)
④获取执行轨迹x_i函数调用树所有节点的关键词列表并集$KW(x_i)=\cup_{1\le j\le |x_i|} KW(x_i)_{mcj}$
⑤利用tf-idf模型获取执行轨迹x_i排序最高的k个关键词列表$KW(x_i, k)$
⑥更新执行轨迹集合X的词汇表$V=V\cup KW(x_i, k)$
⑦更新执行轨迹索引i=i+1
i<|X|+1?
是
否
⑧输出执行轨迹集合X的词汇表V

图3

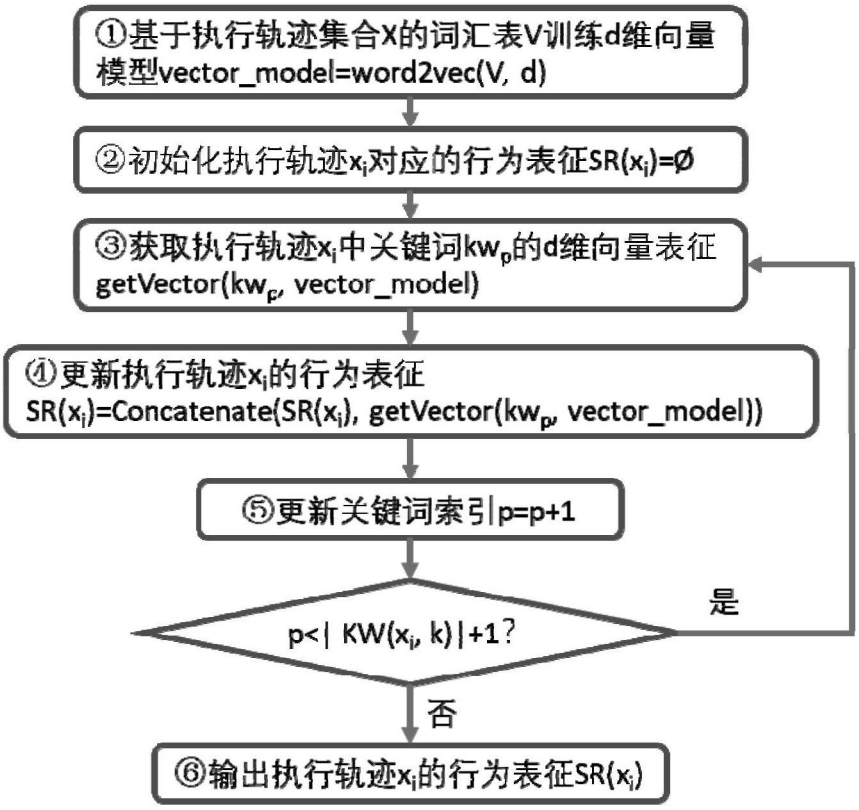

①基于执行轨迹集合X的词汇表V训练d维向量模型vector_model=word2vec(V, d)
②初始化执行轨迹x_i对应的行为表征$SR(x_i)=\emptyset$
③获取执行轨迹x_i中关键词kw_p的d维向量表征 getVector(kw_p, vector_model)
④更新执行轨迹x_i的行为表征 $SR(x_i)$=Concatenate($SR(x_i)$, getVector(kw_p, vector_model))
⑤更新关键词索引p=p+1
$p<|KW(x_i, k)|+1$?
是
否
⑥输出执行轨迹x_i的行为表征$SR(x_i)$

图4

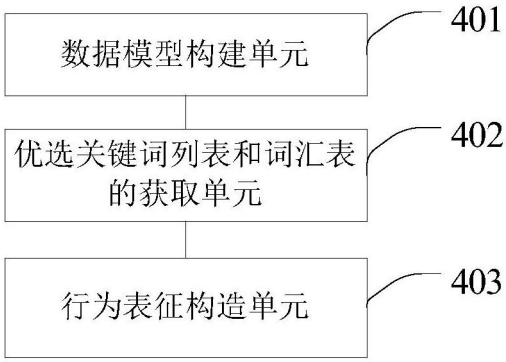

数据模型构建单元
401
优选关键词列表和词汇表的获取单元
402
行为表征构造单元
403

图5

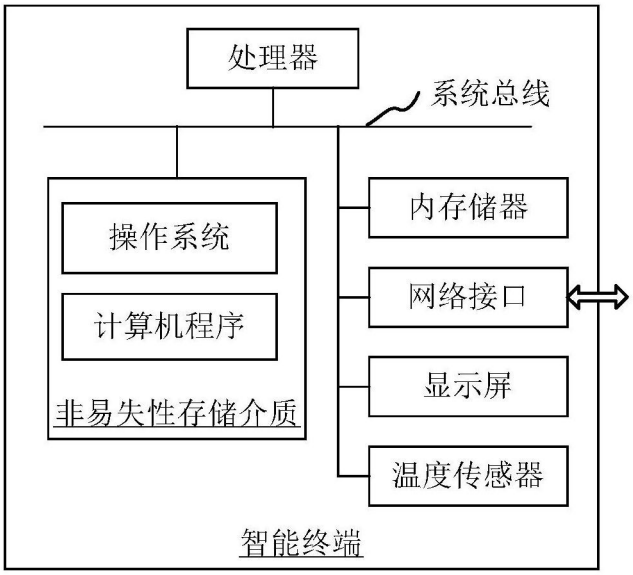


图6